Thèse de Sciences

Université Paris Descartes

Sciences de la vie et de la Santé

Discipline : Génétique
Ecole Doctorale Gc2iD

Guillaume Huguet

Identification de facteurs génétiques impliqués dans les troubles du spectre autistique et de la dyslexie

Soutenue publiquement le 26 novembre 2013

Devant le jury composé de :

Président du jury : Mr. Jan Sap, PU Université Paris Diderot

Rapporteur : Mme Frédérique Bonnet-Brilhault, PUPH ; Centre universitaire de pédopsychiatrie de Tours

Rapporteur : Mr. Damien Sanlaville, PUPH; Université Claude Bernard Lyon 1

Examinateur : Mr. Daniel Moreno de Luca, PhD; Université de Yale

Examinateur : Mme Fiona Francis; PhD, INSERM, Institut du Fer à Moulin

Directeur de thèse : Mr. Thomas Bourgeron, PU Université Paris Diderot

A tous ceux qui m'ont éclairé au cours du chemin.

« Dans les sciences, le chemin est plus important que le but.

Les sciences n'ont pas de fin. »

de Erwin Chargaff

« La route ? Là où on va, on n'a pas besoin de route !»

de Dr Emmett Brown (Back to the futur I)

Remerciements :

Je tiens à remercier mon directeur de thèse Thomas Bourgeron, qui a su m'encadrer, me stimuler et me donner la possibilité de travailler dans son laboratoire où j'ai pu trouver un environnement scientifique qui me convenait. J'ai beaucoup appris à son contact et souhaite encore longtemps travailler avec lui. Plus qu'un directeur de thèse, il a su être mon mentor au cours de ces 4 années.

Je voudrais aussi remercier certains de mes professeurs, qui m'ont aidé dans cette ascension grâce à leur bienveillance, leur foi en moi et leurs encouragements. En commençant par Mme Louis, directrice de l'école primaire des Vieille Vigne de la Ferté-Alais puis Mme Béchéqui (Prof d'anglais), Mme Brothier (Prof de S.V.T.) et Mme Prigent (Prof d'Arts plastiques) du collège Léonard de Vinci. Ensuite ce fut Mme Galare (Prof d'Histoire/Géographie) et Mr Fournier (Prof d'initiation au Sciences de l'Ingénieur) du Lycée Geoffroy Saint Hilaire qui ont su me soutenir. Enfin, merci à Catherine Alcaide, Claude Bazin et Fabien Fauchereau, de l'université Paris 7.

Au sein du laboratoire, je remercie :

Nathalie Lemiére dit la « Viking », pour sa bonne humeur et son aide tout au long de ces années de thèse. C'est une personne qui a une main de fer dans un gant de velours. Impossible de ne pas évoquer son goût prononcé pour l'orthographe, même si j'ai essayé de l'initier au langage SMS. Merci aussi à Anne-Marie et Corinne pour leurs sourires et leur gentillesse au quotidien.

Béatrice Regnault et Laure Lemée pour l'excellent travail qu'elles ont produit et leur expertise sur la technologie Illumina qui a été sans faille. Puis je voudrais rajouté pour Laure, qu'il faut pas que tu changes, on est tous fou de son coté ROCK.

Laura que je remercie d'être elle-même, cette Princesse Parisienne, qui est généreuse et toujours à la pointe de l'élégance, mais aussi pour son aide dans la correction de la thèse. Après promis, on lance notre série « the planning ».
Vu qu'une princesse n'est jamais seule, je veux aussi remercier Anna B. pour son courage dans la quête des fautes d'orthographe et pour sa gentillesse, ainsi que aussi Marion pour sa ténacité et son aide dans la recherche des mutations. Pour finir, je remercie Ophélie qui par sa fraicheur a su apporter un nouveau visage au laboratoire.
En revanche, il n'y a pas de prince charmant à remercier mais des Chevaliers sans peur et sans reproche, au nombre de 3 : Alex et Julien ont su m'aider dans les projets et dans les problèmes Informatiques en s'infiltrant dans la Matrix pour la débugger.
Raphael Kech le Padawan, qui a été d'une grande aide par son travail au cours de son stage de M2, et est même, avec mon enseignement, passé du côté obscur de la force. Il est et sera toujours un excellent étudiant bourré de talents.

J'aimerai insister sur mon frère d'arme, Alex, qui sans son aide précieuse, je n'aurais pu avancer aussi vite et appris autant dans le débuggage de la Matrix. Je suis content de le compter dans mes amis les plus proches.

Passons maintenant à Claire, Cécile, Elodie et Oriane pour toutes ces années de rire, de soutien et d'entraide que l'on a vécu ensemble, dans le périple qu'est la thèse. Je vous remercie du font du cœur pour tout ce que l'on a pu partager ensemble.

Je tiens à remercier « the team psy » de choc avec Anna M. , Anita, Aline, Richard et Frédérique, qui maintiennent l'équilibre mental du laboratoire. Merci également pour leur aide, leur bonne humeur et le vent energisant qu'ils soufflent à chaque fois qu'ils sont présents. Petit message spécial pour Anna et toutes ces musiques dans la tête que l'on s'est inceptionné et Richard pour son rôle de grand frère au sein du laboratoire.

Je souhaite enfin remercier Hany et Isabelle pour leur contribution, Roberto pour savoir faire tourner les cerveaux et son aide dans les arcanes obscurs des statistiques.
Pour l'ancien du laboratoire Jean-Pierre, merci pour ton savoir intarissable de tu transmets et les nombreuses conversations que l'on a pu avoir.

Je tiens à remercier tout particulièrement mes parents et ma sœur, pour leur soutien et patience tout le long de mes études, et bien sûr pour leur bienveillance et l'aide qu'ils ont su me donner tout au long de ma vie. Sans eux je ne sais pas où j'en serais aujourd'hui. Merci à Jimmy, mon meilleur ami depuis 18 ans, qui a vu l'évolution de mes projets jusqu'à leur aboutissement, qui a toujours sur m'aider pour décompresser mais aussi me montrer le bon chemin quand je manquais de perspectives dans ma vie. « A la vie à la mort » comme on se l'ait toujours dit.

Je tiens aussi à faire un hommage à une amie très chère avec qui j'ai fait le pari un jour pour que l'on se soutienne l'un l'autre dans nos études et qui n'est plus là pour le voir, merci d'avoir était là Kelly, J'y suis arrivé !

Je n'oublie pas tous mes amis qui m'ont permis de vivre en plus de ma thèse des moments inoubliables de rire et joie, ce qui m'a aidé à tenir bon tout au long de ces épreuves grâce à Charlène, Brice, Maylis, Cédric, Mimi, Mélo et Esther avec les mardis CINE, les mercredi DVD, les Sorties Musée les weekend Arrosés et les soirées pyjamas.

Je garde le meilleur pour la fin, je remercie mon cœur, avec qui je partage ma vie depuis plus de 2 ans, qui a su me supporter dans mes excentricités et mes lubies. Tu as vécu cette thèse avec moi avec un soutien sans faille, je tiens à te remercier, je t'aime.

Résumé :

Les troubles du spectre autistique (TSA) touchent approximativement 1% de la population général. Ces troubles se caractérisent par un déficit de la communication sociale et des comportements stéréotypés et restreints. Plusieurs gènes impliqués dans le déterminisme des TSA ont été identifiés comme par exemple les gènes *NLGN3-4X, NRXN1-3 et SHANK1-3*. Au cours des années précédentes, les TSA ont été considérées comme un ensemble complexe de troubles monogéniques. Cependant, les études récentes du génome complet suggère la présence de gènes modificateurs (« multiple hits model »).

La dyslexie est caractérisée par un trouble dans l'apprentissage de la lecture et l'écriture qui touche 5-15% de la population général. Les facteurs génétiques impliqués restent pour l'instant inconnu car seulement des gènes ou loci candidats ont été identifié.

Mon projet de thèse avait pour objectif de poursuivre l'identification des facteurs génétiques impliqués dans les TSA et de découvrir un premier facteur génétique pour la dyslexie. Pour cela, deux types de populations ont été étudiés : d'une part des patients atteints de TSA (N>600) provenant de France, de Suède et des Iles Faroe, d'autre part des patients atteints de dyslexie (N>200) provenant de France en particulier un famille de 11 personnes atteintes sur 3 générations. J'ai utilisé à la fois la technologie des puces à ADN Illumina (600 K et 5M) et le séquençage complet du génome humain pour effectuer des analyses de liaison et d'association.

Pour les TSA, grâce aux analyses de CNVs, j'ai pu identifié des gènes candidats pour l'autisme et confirmé l'association de plusieurs gènes synaptiques avec l'autisme. En particulier, l'étude d'une population de 30 patients des îles Faroe a pu confirmé l'implication des gènes *NLGN1* et *NRXN1* dans l'autisme et identifié un nouveau gène candidat *IQSEC3*. En parallèle, j'ai exploré *PRRT2* localisé en 16p11.2. *PRRT2* code pour un membre du complexe SNARE synaptique qui permet la libération des vésicules synaptiques. Je n'ai pas pu mettre en évidence d'association avec les TSA, mais j'ai montré que ce gène important pour certaines maladies neurologiques était sous pression de sélection différente selon les populations.

Pour la dyslexie, j'ai effectué une analyse de liaison (méthode des lod-scores) pour une grande famille de 11 individus atteints sur trois générations. Cette étude a permis d'identifier *CNTNAP2* comme un gène de vulnérabilité à la dyslexie. Cette découverte est importante car ce même gène est aussi associé aux TSA. Par contre aucune des 20 variations rares découvertes par le séquençage complet du génome n'est localisée dans les parties codantes du gène. Plusieurs variations localisées dans des régions régulatrices sont candidates.

En conclusion, les résultats de ma thèse ont permis d'identifier des gènes candidats pour les TSA, de confirmer le rôle des gènes synaptiques dans ce trouble, de montrer pour la première fois grâce à une analyse de liaison le rôle de *CNTNAP2* dans la dyslexie.

Sommaire :

Introduction

1. Hétérogénéité clinique et biologique des troubles du spectre autistique

Les Troubles du Spectre Autistique (TSA) sont très hétérogènes tant sur le plan clinique qu'étiologique. Les premières descriptions clinques de ces troubles remontent au début des années 40 où ces troubles ont été décrits de façon indépendente par le pédopsychiatre américain d'origine austro-hongroise, Leo Kanner et par le psychiatre autrichien Hans Asperger. De nos jours, la description des TSA est toujours en évolution, liée à de nouveaux instruments d'évaluation clinique mais aussi à l'évolution de la recherche qui permettent ainsi d'explorer de nouveaux aspects encore inconnus de l'autisme.

1.1 Autisme et TSA

La triade autistique

Les travaux effectués dans les années 40 par Leo Kanner (en 1943) aux Etats-Unis, puis par Hans Asperger (en 1944) en Autriche, ont permis pour la première fois de décrire la clinique des TSA. Leo Kanner, sur la base de 11 cas (8 garçons et 3 filles, âgés de 2 à 11 ans), décrit des incapacités innées à créer des contacts affectifs avec autrui ainsi qu'un "désir obsessionnel d'immuabilité" Kanner (1943). Kanner indique entre autre que le développement de ces enfants est différent des patients schizophrènes, car l'isolement est présent très précocement, ce qui n'est pas le cas dans la schizophrénie. Indépendemment, Hans Asperger décrit, chez 4 garçons (âgés de 6 à 11 ans) (Asperger 1944) présentant des capacités intellectuelles normales (voire des talents particuliers),

une «psychopathie autistique» caractérisée par «un manque d'empathie, une faible capacité à se faire des amis, une conversation unidirectionnelle, une forte préoccupation vers des intérêts spéciaux, et des mouvements maladroits». Il les définit par un «dérèglement fondamental […] entraînant des difficultés d'intégration sociale considérables et très typiques», pouvant être compensées avec «une originalité de la pensée et du vécu qui peut mener, par la suite, à des accomplissements exceptionnels» (Asperger 1944). Les travaux d'Asperger, écrits en allemand, sont restés méconnus jusqu'en 1981, date à laquelle la psychiatre britannique Lorna Wing, les remit au goût du jour. Dès lors, sur la base des 4 cas présentés par Asperger et de trente-quatre autres cas issus de sa propre clinique, elle propose une nouvelle définition du syndrome d'Asperger, adaptée aux patients présentant cette clinique particulière(Wing 1981).

Ainsi en 1981, Lorna Wing caractérise la clinique des TSA et définit le concept de « triade autistique», qui s'observe dans la petite enfance, parfois après une phase de développement normal (notion de régression) et regroupant les symptomes suivants:

- troubles des interactions sociales,
- troubles de la communication,
- comportements stéréotypés et intérêts restreints.

Cette triade induit un handicap fonctionnel de l'individu. Même si le syndrome d'Asperger ne comprend pas de troubles des capacités intellectuelles, l'autisme est souvent associé à une déficience intellectuelle mais aussi à d'autres comorbidités. Cette hétérogénéité clinique définie la notion de spectre autistique.

Le spectre autistique

Actuellement les TSA s'inscrivent dans les troubles neuro-développementaux, qui incluent notamment les troubles du développement intellectuel, le trouble de déficit de l'attention/hyperactivité (TDAH), ou encore certains troubles moteurs comme les maladies de tics chroniques.

Néanmoins il existe différentes manières pour catégoriser les patients atteints d'un TSA. L'un d'eux est le DSM-IV (The Diagnostic and Statistical Manual of Mental Disorders (DSM)). publié par l'American Psychiatric Association, il fournit un langage commun et des critères normalisés pour la classification des troubles mentaux. Créé en 1952, le DSM est régulièrement mis à jour par « The International Statistical Classification of Diseases and Related Health Problems » (ICD) et largement utilisé à travers le Monde, pour la clinique et la recherche.

Le DSM-IV (1994) catégorisait l'autisme dans les troubles envahissants du développement (TED), et les répartit en cinq catégories :

- Le trouble autistique, regroupant les troubles des interactions sociales, de la communication, les comportements stéréotypés et les intérêts restreints.

- Le syndrome d'Asperger, qui se distingue de l'autisme par l'absence de retard de langage et de déficit cognitif.

- Le trouble désintégratif de l'enfance, caractérisé par les mêmes critères que l'autisme mais ne fait son apparition qu'après l'âge de 2 ans (développement normal jusque l'âge de 2 ans, suivi d'une perte progressive des acquisitions fonctionnelles cognitives et psychomotrices).

- Le trouble envahissant du développement non spécifié (TED-NOS), qui regroupe les autistes atypiques (défini par la symptomatologie et/ou l'âge d'apparition).

- Le syndrome de Rett, présent uniquement chez les filles et dont la cause génétique est connue. Ce syndrome rare a des similitudes avec le trouble désintégratif de l'enfance, puisque celui-ci aussi survient après une phase de développement normal mais est caractérisé par des déficiences sévères multiples, motrices et cognitives.

Une nouvelle version de cette catégorisation, le DSM-5.0, isole le syndrome de Rett du reste des 4 autres catégories, qu'il regroupe dans une catégorie unique, le « trouble du spectre autistique ». Au niveau de la triade celles-ci se voient réduites à deux : déficits de la communication sociale d'une part, et comportements répétitifs et intérêts restreints d'autre part, en spécifiant une gamme de sévérité par critères (American Psychiatric Association, 2012).

D'autres nomenclatures existent, comme le concept d'ESSENCE de Christopher Gillberg (ESSENCE pour « Early Symptomatic Syndromes Eliciting Neurodevelopmental Clinical Examinations »)(Gillberg 2010). Ce concept permet d'avoir une définition plus complète et moins catégorielle de l'hétérogénéité clinique observée chez les enfants avant l'âge de 5 ans. Ainsi il définit les symptômes observés dans le trouble du spectre autistique avec une combinaison d'autres signes cliniques (comorbidités).

Les principaux symptômes évalués par ESSENCE dans le développement général sont les suivants (figure 1) :

- Troubles de la communication sociale
- Troubles du langage
- Troubles de la coordination motrice
- Troubles de l'attention
- Hyperactivité motrice
- Troubles du comportement restreint et stéréotypé
- Troubles du sommeil

Il est vrai que l'on observe rarement les symptômes de la TSA de manière isolée ; ces derniers coexistent souvent avec d'autres troubles psychiatriques et médicaux (tels que la déficience intellectuelle, l'épilepsie, les problèmes de coordination, les troubles de l'attention, les tics, l'anxiété, les troubles du sommeil et les problèmes gastro-intestinaux). De plus, la variabilité phénoménologique est considérable au sein de chaque symptôme clinique en termes d'intensité. Par exemple, au niveau social, cela peut aller de la simple maladresse à un retrait social majeur. De même, les troubles du langage peuvent aller d'une absence totale de communication à un langage développé mais marqué par un déficit de la pragmatisme.

1.2 Diagnostic et instruments d'évaluation

Epidémiologie

La prévalence actuelle est estimée à 1% (Fombonne, 2009) pour les TSA. Cependant cette prévalence a très fortement augmenté lors de ces 30dernières années. (voir plus bas). Les résultats des études épidémiologiques avant les années 2000 montrent que ~1/3 des personnes avec TSA n'avaient pas de déficience intellectuelle alors que pour les études après les années 2000 ,rapportent que ~2/3 des patients ayant un TSA n'ont pas de déficience intellectuelle (Elsabbagh, Divan et al. 2012).

Les TSA présentent un déséquilibre de répartition de la fréquence des garçons atteints par rapport aux filles, avec un sex-ratio moyen de 4:1, et cette différence augmente d'autant plus que les TSA n'est pas associé à une déficience intellectuelle. Par exemple chez les patients ayant un syndrome d'Asperger, le ratio est de 6:1, à l'oppposé des patients ayant un TSA associés à une déficience intellectuelle, où le ratio est de 2:1 (Fombonne 2003). Plusieurs hypothèses tentent d'expliquer ce déséquilibre du sex-ratio ; par exemple l'implication de facteurs hormonaux et/ou de facteurs génétiques liés aux chromosomes sexuels.

De nombreux constats font état d'une augmentation dans le temps de la prévalence de ces troubles. Cela pourrait s'expliquer par l'amélioration de la méthodologie des études épidémiologiques mais aussi par l'élargissement des critères diagnostiques du spectre autistique. D'un point de vue biologique, l'âge des parents à la naissance des enfants (en augmentation progressive), facteur de risque avéré de TSA (Kolevzon, Gross et al. 2007), pourrait expliquer l'augmentation du nombre de cas diagnostiqués. Toutefois, l'impact des facteurs environnementaux reste à mieux caractériser.

Diagnostic

Le diagnostic clinique des TSA est basé à la fois sur un entretien avec les parents et sur une observation clinique de l'enfant selon des instruments de diagnostic standardisés comme l'ADI et l'ADOS, qui permettent de diagnostiquer la présence ou non d'un TSA.

- L'ADI-R (Autism Diagnostic Interview-Revised) (Lord, Rutter et al. 1994), est un entretien composé d'une centaine de questions à poser aux parents.

- L'ADOS (Autism Diagnostic Observation Schedule) (Lord, Risi et al. 2000), regroupe différents séquences de jeux adaptés au niveau de langage et de développement de l'enfant, permettant de juger de l'interaction

Malgré les qualités métrologiques de l'ADI et l'ADOS, persite un manque de résolution quantitative et qualitative des symptômes autistiques et de fiabilité inter-cotateurs. Une récente étude montre que les critères de classification des sous-types de TSA, avec ADI et ADOS, varient significativement d'un centre à un autre (Lord, Petkova et al. 2012). D'autres outils de screeening sont parfois utilisés permettant une évaluation quantitative des traits autistiques, comme Autism Spectrum Quotient(AQ) ou la Social Responsivnes Scale (Baron-Cohen, Wheelwright et al. 2001). Par exemple, la distribution du score de l'AQ est représentée par une gaussienne dans la population générale qui, quand il est faible, définit une personne avec peu ou pas de traits autistiques, et, lorsqu'il augmente, montre une sévérité croissante. La moyenne de ce score dans la population générale est de 16,4, et chez les patients ayant un TSA, d'au moins 32. Il est aussi observé que ce score est en moyenne plus faible chez les femmes (Berkel, Tang et al. 2011) que les hommes (Bittel, Kibiryeva et al. 2006). Néanmoins 2%

de la population générale atteint des scores supérieurs ou égaux à celui observé chez les autistes, sans pour autant présenter de diagnostic de TSA. Par ailleurs, lorsque l'on fait passer l'AQ chez les apparentés de patients ayant un TSA (Wheelwright, Auyeung et al. 2010), des scores plus élevés en moyenne chez eux que dans la population générale sont relevés. Cela a abouti au concept de « phénotype autistique élargi » (broader autism phenotype) (Bailey, Palferman et al. 1998).

Les autres outils nécessaires à l'investigation clinique de ces patients sont les examens neuropédiatrique, associés à des examens visuels, auditifs, métaboliques et génétiques (en particulier, X Fragile, caryotype en l'absence de puce (CGH ou SNPs) à haut débit,) (Haute Autorité de Santé, 2010).

Comorbidités et signes associés

Les TSA sont rarement isolés, d'autres troubles alors appelés comorbidités sont fréquemment associés (cf. Tableau page suivante). Même si ces comorbidités ne font pas partie de la définition de la triade autistique au sens du DSM-IV, elles jouent un rôle majeur dans le handicap, parfois même plus que les symptômes autistiques, eux même. Par exemple, la déficience intellectuelle, qui selon le DSM-IV se répartie en 4 catégories selon son degré de sévérité (déficience intellectuelle léger (50<QI<70), modéré (35<QI<50), sévère (20<QI<35), profond (QI<20)) entraine un handicap fonctionnel sévère et supplémentaire. Cependant, cette classification du déficit intellectuel n'est pas toujours applicable dans les TSA, car les capacités cognitives des patients sont hétérogènes. Souvent les capacités verbales sont plus atteintes, que les capacités de performances, mais cela peut être parfois l'inverse.

Trouble	Commentaire	Observation et prévalence
Déficience intellectuelle	Prévalence varie de manière importante selon les études	70% dans l'autisme sans le syndrome Asperger (Chakrabarti and Fombonne 2005), (Fombonne 2003) (note 1)
Dysmorphies	Oriente souvent le diagnostic vers un syndrome génétique	16% des patients présentent des dysmorphies ; celles-ci sont associées à un QI plus faible et à de l'épilepsie. (Miles, Takahashi et al. 2005) (N=260 cas)
Anomalies du périmètre crânien	Une augmentation du périmètre crânien associée à une augmentation du volume cérébral chez les cas, (La méta-analyse de Redcay et al. (2005)(Redcay and Courchesne 2005)) (N=531 cas) (note 2)	Volume cérébral supérieur à la moyenne dans 90% des cas, avec une macrocéphalie (>+2DS dans 37% des cas) ceci est associé à une augmentation de la matière blanche et grise corticale, et, au niveau du cervelet, à une augmentation de la matière blanche mais diminution de la matière grise (augmentation du nombre et/ou de la taille des cellules débat) (Courchesne, Karns et al. 2001) (N=30 cas males- entre 2-4 ans) La microcéphalie est observée chez une minorité (5 à 15%) de cas et est corrélée à la présence d'une autre pathologie.
Épilepsie	plus fréquente chez les filles surtout lorsqu'elles ont une déficience intellectuelle ((Amiet, Gourfinkel-An et al. 2008))	20-25% (Canitano 2007)

Tableau des co-mobidités partie 1.

Note 1- Il existe de nombreux outils pour évaluer le fonctionnement intellectuel : les échelles standardisées de mesure du quotient intellectuel (QI) qui distinguent classiquement le QI Verbal et le QI Performance, les matrices progressives de Raven (test de logique permettant d'évaluer l'intelligence indépendamment des capacités verbales), ou encore le profil psycho-éducatif (psychoeducative profile PEP-R).

Note 2- Cette augmentation du périmètre cranien est en lien avec une anomalie de la cinétique de croissance au début de la petite enfance (<24 mois), alors que le périmètre cranien des enfants ayant un TSA est normal à la naissance.

Trouble	Commentaire	Observation et prévalence
Troubles moteurs	La fréquence diminue chez les cas plus âgés, ce qui suggère que la motricité tend à se normaliser	51% hypotonie (sur N=154 cas) (Ming, Brimacombe et al. 2007)
Déficience auditive		9,5% des patients, dont 3,5% de surdité (sur N=199 cas) (Rosenhall, Nordin et al. 1999) (note 3)
Anomalies sensorielles	Hyposensibilité ou hypersensibilité aux stimuli auditifs, visuels, tactiles, olfactifs, gustatifs	(Baron-Cohen, Ashwin et al. 2009)
Les troubles psychiatriques	Prévalences varient selon les études,on observe des troubles anxieux, de déficit de l'attention/hyperactivité (TDAH), de l'humeur, la phobie sociale, d'oppositionnel, de type agressivité et auto-agressivité	- TDAH 55%, de phobie spécifique 44%, de trouble obsessionnel compulsif (TOC) 43%, de dépression 25% (sur N=109 cas) (Leyfer, Folstein et al. 2006) - la phobie sociale 29%, le TDAH 28% et le trouble oppositionnel 28% (sur N=112 cas) (Simonoff, Pickles et al. 2008) - troubles du comportement de type agressivité et/ou auto-agressivité 22% (Hartley, Sikora et al. 2008)
Les troubles du sommeil	Insomnie chronique	Insomnie chronique est 10 fois plus élevée et les troubles du sommeil persistent avec seulement 8% de rémission (vs 52% chez les témoins). (N=3700 dont 28 avec TSA) (Sivertsen, Posserud et al. 2012)
Les troubles alimentaires	Sélectivité alimentaire	25% des patients avec TSA (note 4)
Constipation		34% (sur N=121 cas) (Ibrahim, Voigt et al. 2009)

Tableau des co-mobidité partie 2.

Note 3- En revanche, la suspicion d'une association entre TSA et maladie cœliaque, à l'origine du régime sans gluten et sans caséine parfois proposé aux patients, n'a pas été confirmée (voir la synthèse de consensus de la Haute Autorité de Santé, 2010).

Note 4- Néanmoins, ce chiffre devrait être réexaminé avec la pratique systématique d'une audiométrie devant un retard d'acquisitions.

1.3 Prise en charge

Le choix de la prise en charge est comme le syndrome, tout aussi diversifié. Ce choix est tout de même souvent déterminé par le niveau intellectuel du patient. Pour les individus avec un QI normal ou supérieur à la moyenne, comme les patients avec autisme de haut niveau ou syndrome d'Asperger, une scolarité normale est possible.

Actuellement, à l'échelle internationale, de nombreux organismes ont pour volonté de mettre en place une prise en charge pour les patients avec TSA. Les méthodes développées se basent sur la stimulation de l'enfant très précocement pour permettre d'améliorer les performances cognitives, le langage et les capacités d'interaction sociale. Malgré des résultats prometteurs, ces méthodes, comme la méthode ABA (Applied Behaviour Analysis) ou la méthode TEACCH (Treatment and Education of Autistic and related Communication handicaped CHildren), restent malheureusement peu développées en France. Par exemple, basée sur l'approche comportementale, la Thérapie d'Echange et de Développement, développée à Tours, a pour but, lors de séances de jeu ou de plaisir partagé, de développer au cours du temps les fonctions de base du développement d'une communication, à savoir les fonctions d'imitation, d'attention et de perception chez le patient (*Barthélémy C., Hameury, L et Lelord G. La thérapie d'échange et de développement dans l'autisme de l'enfant. Expansion Scientifique Française éd. Paris 1998, 302)*

Les traitements médicamenteux sont actuellement pas ou peu efficaces sur le syndrome autistique (interaction sociale) mais ciblent plutôt les comorbidités. Ce sont souvent des médicaments psychotropes développés dans d'autres pathologies comme les antipsychotiques (risperidone, aripiprazole) pour les troubles du comportement, le méthylphénidate pour l'hyperactivité, et les inhibiteurs sélectifs de la recapture de la

sérotonine (fluoxétine, citalopram) pour les stéréotypies et les comportements obsessionnels (McPheeters, Warren et al. 2011) et la mélatonine pour le traitement des troubles du sommeil.

De nouvelles perspectives de traitement sont développées, comme avec le bumétanide dans une étude-pilote, qui permettrait de restaurer l'homéostasie des synapses inhibitrices du système GABAergique chez les patients où celui-ci est altéré, en diminuant les concentrations intracellulaires d'ions chlorure (Lemonnier and Ben-Ari 2010). D'autres médicaments sont aussi trés prometteurs, comme le fenobam dans le syndrome de l'X fragile (Berry-Kravis, Hessl et al. 2009) mais aussi l'ocytocine, qui semble, chez des patients autistes de haut niveau, améliorer le comportement social (Andari, Duhamel et al. 2010).

1.4 Facteurs genetiques impliqués dans l'autisme

Il y existe une forte composante génétique à la prédisposition aux TSA, comme l'indique le risque de récurrence dans les études de familles et de jumeaux (Bailey, Le Couteur et al. 1995) (Ronald, Larsson et al. 2010) (Hallmayer, Cleveland et al. 2011) (Lundstrom, Chang et al. 2012). Le taux de récurrence dans les familles ayant un enfant autiste a été initialement estimé à 5%, contre 1% dans la population générale, mais des études récentes observant les signes précoces de TSA au sein de fratries des individus atteints de TSA révèlent un taux de récurrence plus élevé (jusqu'à 20 %) (Ozonoff, Young et al. 2011). Les premières études de jumeaux décrivent les TSA comme les troubles neuropsychiatriques les plus fortement influencés par la génétique, avec des taux de concordance atteignant 82-92% chez les jumeaux monozygotes par rapport à 1-10% chez les jumeaux dizygotes (Bailey, Le Couteur et al. 1995), (Folstein and Rutter 1977).

Cependant des études plus récentes ont indiqué que la concordance chez les jumeaux dizygotes pourrait être plus élevée (> 20%) (Hallmayer, Cleveland et al. 2011). Dans toutes les études, la concordance chez les jumeaux monozygotes n'est pas complète, ce qui indique que des facteurs épigénétiques, stochastiques et / ou environnementaux jouent un rôle (Lichtenstein, Carlstrom et al. 2010), (Lundstrom, Chang et al. 2012), (Ronald, Larsson et al. 2010). Comme illustré dans le paysage épigénétique de Conrad Waddington (Waddington 1942), le phénotype TSA peut résulter d'un réseau de régulation complexe impliquant des facteurs génétiques, épigénétiques et environnementaux ainsi que des fluctuations stochastiques, que l'on peut représenté par une balle qui serait guidée du point le élevé au point le plus bas (figure 1b). Pour chaque individu avec un génotype donné, la réponse à l'environnement peut être différente. Woltereck (Woltereck 1909) introduit les concepts de normes de réaction et de plasticité phénotypique au début du XXe siècle permettant de tenir compte du fait que les organismes peuvent modifier leurs caractéristiques en réponse à l'environnement (figure 1C). En accord avec cette idée, différents génotypes augmentant le risque de TSA, pourraient avoir une « norme de réaction » différente à travers une gamme d'environnements. Ainsi, nous espérons qu'une meilleur connaissance des causes génétiques des TSA devrait nous aider à identifier le meilleur environnement pour un individu donné.

Les premières causes génétiques des TSA ont été identifiées pour les troubles monogéniques tels que le syndrome de l'X fragile et le syndrome de Rett (Pieretti, Zhang et al. 1991, Amir, Van den Veyver et al. 1999) ainsi que dans les familles porteuses d'anomalies chromosomiques (Vorstman, Staal et al. 2006). Plus tard, des mutations délétères ont été identifiées au sein de plusieurs gènes candidats, tels que ceux codant pour les neuroligines, neurexines et SHANKS, qui ont tous une fonction à la synapse et

qui sont des acteurs clés dans la susceptibilité aux TSA, y compris chez les personnes ayant un TSA mais pas de déficience intellectuelle, tel que le syndrome d'Asperger (Durand, Betancur et al. 2007), (Jamain, Quach et al. 2003), (Szatmari, Paterson et al. 2007). Les études par « puces à ADN » ou « micro-array » ont pu révéler des délétions et duplications, appelées variants du nombre de copies (CNV), touchant de nombreux loci, dont 5-15% sont des événements *de novo* chez les patients atteints de TSA (Christian, Brune et al. 2008), (Cooper, Coe et al. 2011), (Gilman, Iossifov et al. 2011), (Glessner, Wang et al. 2009), (Jacquemont, Sanlaville et al. 2006), (Marshall, Noor et al. 2008), (Pinto, Pagnamenta et al. 2010), (Sanders, Ercan-Sencicek et al. 2011), (Sebat, Lakshmi et al. 2007), (Szatmari, Paterson et al. 2007). Plus récemment, les séquençages d'exomes ont été utilisés pour détecter des mutations *de novo* délétères chez 3,6 à 8,8% des personnes atteintes de TSA (Kong, Frigge et al. 2012), (Neale, Kou et al. 2012), (O'Roak, Deriziotis et al. 2011), (O'Roak, Vives et al. 2012), (Sanders, Murtha et al. 2012).

En résumé, au-delà de la définition des TSA se trouve un degré extrême d'hétérogénéité clinique, allant de déficiences légères à profondes. Des études génétiques moléculaires ont révélé le paysage génétique très hétérogène des TSA, avec différents types d'anomalies situées sur presque tous les chromosomes avec différents niveaux de pénétrance (Abrahams and Geschwind 2008), (Devlin and Scherer 2012), (Freitag 2007) (Huguet, Ey et al. 2013).

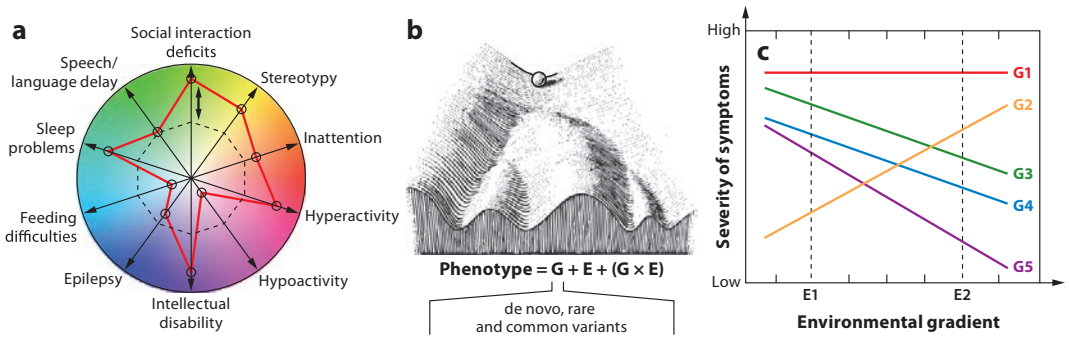


Figure 1 : Diversité clinique et génétique des TSA. (a) Les symptômes et les comorbidités observés chez les individus atteints de TSA. Les flèches définissent les caractéristiques cliniques des individus , et la ligne rouge définit le niveau de gravité (faible au centre et à la périphérie de élevé). Les couleurs représentent la combinaison des caractéristiques cliniques. (b) Un paysage épigénétique. Comme la bille est guidé vers le point le plus bas, la trajectoire clinique d'un individu atteint de TSA est influencée par des facteurs génétiques (G) , des facteurs environnementaux (E) , et des interactions gènes-environnement (G × E) ; Adapté de Waddington (Waddington 1942). (c) Normes de réaction décrivant la façon dont les individus avec différents génotypes répondent à différents environnements. L'individu avec le génotype G1 ne change pas indépendamment de l'environnement . L'individu avec le génotype G2 s'aggrave dans l'environnement E2 par rapport à l'environnement E1. Les personnes ayant un génotype G3-G5 s'améliorent dans l'environnement E2 par rapport à l'environnement E1 , mais avec des normes de réaction différentes ; Adapté de Woltereck (Woltereck 1909).

Des variations de nombre de copies (CNV) *de novo* dans les TSA

Des altérations chromosomiques de grande taille (visibles au microscope) ont été rapportées chez moins de 5% des personnes atteintes de TSA. Les plus fréquentes sont situées dans les régions chromosomiques 2q37, 7q11, 15q11-13, 16p11, 22q11.2 et 22q13.3 (Vorstman, Staal et al. 2006). Au cours de la dernière décennie, d'importants progrès dans la génétique des TSA ont émergé à partir d'études de CNV à l'aide de puces à ADN (Alkan, Coe et al. 2011).

La première analyse, qui a utilisé l'hybridation génomique comparative (CGH) sur un petit échantillon de 29 personnes atteintes de TSA, a constaté que 8 d'entre eux (28%) avaient des CNV cliniquement pertinentes (six délétions et deux duplications) (Jacquemont, Sanlaville et al. 2006). Le taux élevé d'altérations génétiques rapportées dans cette étude était très probablement du à l'inclusion d'individus présentant des comorbidités souvent associées à des anomalies génétiques, telles que la dysmorphie faciale.

La deuxième étude principale a utilisé une puce (array) de CGH plus dense et n'a pas été biaisée en faveur des personnes présentant des comorbidités; ici, les auteurs ont constaté des CNVs *de novo* chez 12 des 118 individus atteints de TSA sporadiques (10%), chez 2 des 77 individus atteints de TSA qui ayant également un parent au premier degré atteint (3%), et chez 2 sur 196 témoins (1%) (Sebat, Lakshmi et al. 2007). De nombreuses études ont depuis utilisé la CGH arrays (puce) ou des puces à ADN pour étudier le fardeau des CNVs sur de larges cohortes d'individus atteints de TSA (Christian, Brune et al. 2008), (Cooper, Coe et al. 2011), (Gilman, Iossifov et al. 2011), (Glessner, Wang et al. 2009), (Itsara, Wu et al. 2010), (Jacquemont, Sanlaville et al. 2006), (Marshall, Noor et al. 2008), (Pinto, Pagnamenta et al. 2010), (Sanders, Ercan-Sencicek et al. 2011), (Sebat, Lakshmi et al. 2007, Szatmari, Paterson et al. 2007). À ce

jour, les génotypes de plus de 3800 personnes atteintes de TSA, de 1200 frères et sœurs, de 600 contrôles, ainsi les parents ont été analysés. Dans toutes les études, des CNVs *de novo* semblent être enrichis chez les individus atteints de TSA par rapport à leurs frères et sœurs non atteints et à la population de contrôle. Dans l'ensemble, 6,6% des personnes atteintes de TSA sporadiques avaient au moins un CNV *de novo* rare, comparativement à 4,1% chez les individus atteints de TSA ayant un parent affecté au premier degré, 1,4% chez les frères et les sœurs non atteints et 1,9% chez les témoins (Figure 2a). Notamment, CNVs *de novo* sont plus larges et touchent davantage de gènes chez les individus atteints de TSA par rapport à leurs frères et sœurs sains et des contrôles. Cet enrichissement ne reflète pas un risque plus élevé de parents à produire ces CNVs *de novo*, parce que les frères et les sœurs auraient la même fréquence que les personnes atteintes de TSA, ce qui n'est pas le cas. Comme prévu, les CNVs *de novo* sont plus fréquents chez les personnes atteintes de TSA dans des cas sporadique que chez les individus atteints de TSA qui avaient aussi un parent au premier degré touchée, ce qui indique que la structure de la famille doit être considérée pour estimer la prévalence des CNVs *de novo* dans les TSA.

Deux études de grande envergure ont déterminé l'origine parentale des CNVs *de novo* (Itsara, Wu et al. 2010), (Sanders, Ercan-Sencicek et al. 2011). En combinant les résultats des deux études révèle une tendance des CNVs *de novo* d'origine maternelles (54,5%), mais le modèle n'a pas atteint une signification statistique (Figure 2b). Il est intéressant de noter que, par rapport aux hommes, les femmes atteints de TSA ont plus de CNV *de novo*, et qu'ils touchent davantage de gènes (figure 2c). Enfin, il existe une association significative entre le niveau cognitif, tel que mesuré par le QI et le nombre de gènes affectés par un CNV *de novo* (Girirajan, Rosenfeld et al. 2012), (Sanders, Ercan-Sencicek et al. 2011). Dans une étude réalisée par Girirajan et al. (Girirajan, Rosenfeld et

al. 2012), le QI moyen dépasse le seuil d'une déficience intellectuelle (QI <70) chez les personnes ayant 18 génes ou plus présent dans le CNV *de novo* (figure 2d). En résumé, la charge des CNVs *de novo* et rare est significativement plus élevée chez les individus atteints de TSA par rapport à leurs frères et les sœurs non atteint et la population générale. Les CNVs sont plus nombreux, plus grands et contiennent plus de gènes. La présence d'une dysmorphie faciale et / ou de déficience intellectuelle augmente les chances de détecter un CNV *de novo*. Enfin, les femmes semblent avoir une plus grande tolérance que les hommes d'accumulation des altérations génétiques sans développer un TSA. Pourtant, si les CNVs sont fortement associées aux TSA, surtout dans les cas sporadiques, alors qu'ils ne comptent que pour une petite fraction de personnes atteintes de TSA, et donc d'autres types de mutations doivent jouer un rôle dans la susceptibilité à l'autisme, en particulier dans les cas familiaux .

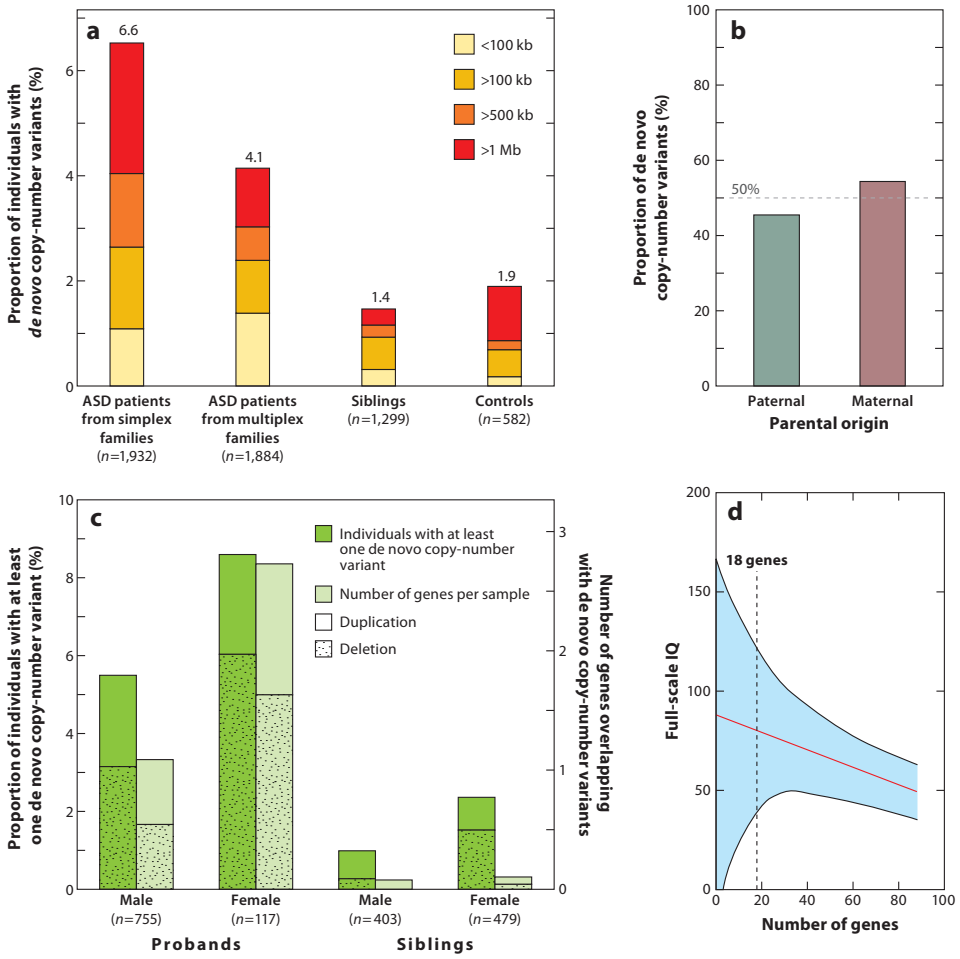


Figure 2 : Le fardeau de CNVs *de novo* dans les TSA. (a) La fréquence et la taille des CNVs *de novo* chez les individus atteints de TSA issus de familles simplex, de familles multiplex, et aussi chez les frères et les sœurs non atteints, et les témoins. Les données proviennent des études suivantes (Marshall, Noor et al. 2008, Itsara, Wu et al. 2010, Pinto, Pagnamenta et al. 2010, Sanders, Ercan-Sencicek et al. 2011) et (Sebat, Lakshmi et al. 2007). (b) L'origine parentale des CNVs *de novo* identifiées chez les individus atteints de TSA. Les données proviennent des études (Itsara, Wu et al. 2010) et (Sanders, Ercan-Sencicek et al. 2011). (c) Proportion de personnes ayant au moins un CNV *de novo*, de même que le nombre de gènes (par RefSeq) présent dans un CNV. Les données sont tirées des études (Sanders, Ercan-Sencicek et al. 2011). (d) Relation entre le QI et le nombre de gènes qui se chevauchent avec le CNVs *de novo*. Les données proviennent des études (Girirajan, Rosenfeld et al. 2012) et (Sanders, Ercan-Sencicek et al. 2011).

Variations *de novo* dans les séquences codantes chez les autistes

Les études sur les variants *de novo* dans les séquences codantes étaient auparavant limitées aux gènes candidats. Les technologies de séquençage de nouvelle génération, telles que le séquençage de l'ensemble de l'exome ou du génome, nous permettent maintenant d'estimer la contribution des mutations *de novo* au sein des séquences codantes dans les TSA. Les six études d'exome publiées à ce jour ont séquencé plus de 1000 personnes atteintes de TSA ainsi que 790 frères et sœurs (O'Roak, Deriziotis et al. 2011, Iossifov, Ronemus et al. 2012, Kong, Frigge et al. 2012, Neale, Kou et al. 2012, Sanders, Murtha et al. 2012). Dans ces études, le taux moyen de variants *de novo* au sein de séquences codantes par personne (y compris les faux-sens, l'épissage, les frameshift et variantes stop-gain) était de 0,86% chez les femmes atteintes de TSA, de 0,73% chez les hommes atteints de TSA et de 0,60% chez les germains non atteints (Figure 3a). Au niveau du groupe, il n'y avait pas de différence statistiquement significative entre les personnes atteintes de TSA, leurs germains non atteints et les contrôles. Ce n'est que lorsque l'analyse se limite aux gènes exprimés dans le cerveau que la différence devient statistiquement significative ($p = 0,001$), avec plus de mutations chez les individus atteints de TSA par rapport aux germains non atteints (Sanders, Murtha et al. 2012). Au niveau individuel, plusieurs mutations causales ont été identifiées, ce qui représente 3.6–8.8% des personnes avec TSA. Iossifov et al. (Iossifov, Ronemus et al. 2012) ont estimé que 65 gènes responsables ont été identifiés par quatre études de séquençage d'exome à grande échelle. Ce calcul est basé sur le nombre de mutations *de novo* délétères identifiées dans les différentes études de séquençage d'exome (59 des 343 proposants dans la référence ((Iossifov, Ronemus et al. 2012), 90 des 175 proposants dans la référence (Neale, Kou et al. 2012), 33 des 209 proposants dans la référence (O'Roak, Vives et al. 2012) et 17 des 238 proposants dans la référence (Sanders, Ercan-

Sencicek et al. 2011)) et d'un taux différentiel double estimée de mutations délétères entre les individus atteints de TSA et leur fratrie. Comme pour les CNVs, les femmes ont tendance à avoir plus de variants *de novo* dans les séquences codantes par rapport aux hommes. Lorsque l'origine parentale des mutations *de novo* à pu être établie, les variants étaient trois fois plus susceptibles de provenir du génome paternel que maternel (Kong, Frigge et al. 2012, O'Roak, Vives et al. 2012) (Figure 3b). L'âge du père à la conception de l'enfant est un facteur qui influence grandement ce taux (Kong, Frigge et al. 2012), avec une augmentation de près de deux mutations par an. Un modèle exponentiel estime que le nombre de mutations paternelles double tous les 16,5 ans (figure 3C).

Les résultat provenant de ces études d'exome n'expliquent malheureusement que peu de cas d'autisme, puisque il n'en sort qu'une poignée de gènes mutés chez chez plusieurs d'individu (*CHD8, KATNAL2, SCN1A, SCN2A, DYRK1A,* et *POGZ*) (Iossifov, Ronemus et al. 2012, Neale, Kou et al. 2012, O'Roak, Vives et al. 2012, Sanders, Murtha et al. 2012). Néanmoins, il est important de garder à l'esprit que les méthodes actuelles utilisées pour les exomes ne saisissent pas tous les exons du génome (principalement lorsqu'ils sont dans des séquences riches en GC) et que de nouvelles technologies de séquençage pourraient être en mesure de détecter des mutations supplémentaires (O'Roak, Vives et al. 2012). La stratification d'individus atteints de TSA en fonction de phénotypes intermédiaires ou de symptômes spécifique permettrait peut être de détecter des gènes avec une prévalence élevée, comme *CHD8*, qui est muté de façon récurrente chez les personnes ayant une macrocéphalie (O'Roak, Vives et al. 2012), et *SHANK3*, qui est muté chez environ 2-3% des personnes présentant une déficience intellectuelle et un TSA (CS Leblond, C. Nava, A. Polge, J. Gauthier, G. Huguet, et al., données non publiées).

En résumé, le fardeau de mutations *de novo* affectant les gènes exprimés dans le cerveau est plus fort chez les individus atteints de TSA par rapport aux témoins. Deux tiers des mutations *de novo* sont d'origine paternelle, et le taux de ces mutations augmente avec l'âge. Actuellement peu de gènes mutés fréquement sont identifiés pour les TSA et les prévisions actuelles estiment que plus de 500-1000 gènes pourraient être impliqués dans les TSA (Iossifov, Ronemus et al. 2012), (Sanders, Murtha et al. 2012).

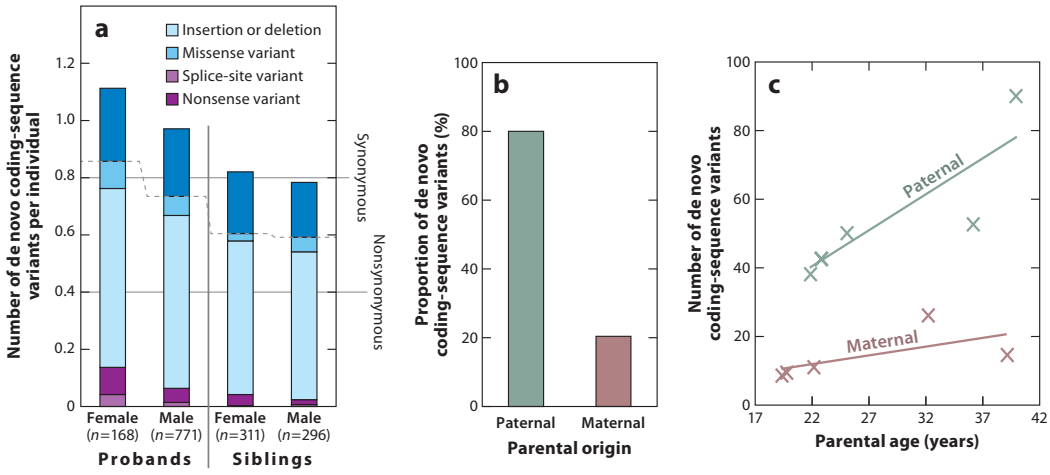


Figure 3 : Le fardeau des variants *de novo* dans les séquences codantes. (a) Distribution des variantes codants *de novo* chez les individus atteints de TSA et les témoins après stratification par sexe. Les données proviennent de références (Iossifov, Ronemus et al. 2012), (Neale, Kou et al. 2012), (O'Roak, Vives et al. 2012) et (Sanders, Ercan-Sencicek et al. 2011). (b) L'origine parentale des variants *de novo* identifiée chez les individus atteints de TSA. Les données proviennent de références (Kong, Frigge et al. 2012) et (O'Roak, Vives et al. 2012). (c) Relation entre l'âge des parents et le nombre de variants *de novo* chez l'enfant. Les données sont tirées de la référence (Kong, Frigge et al. 2012).

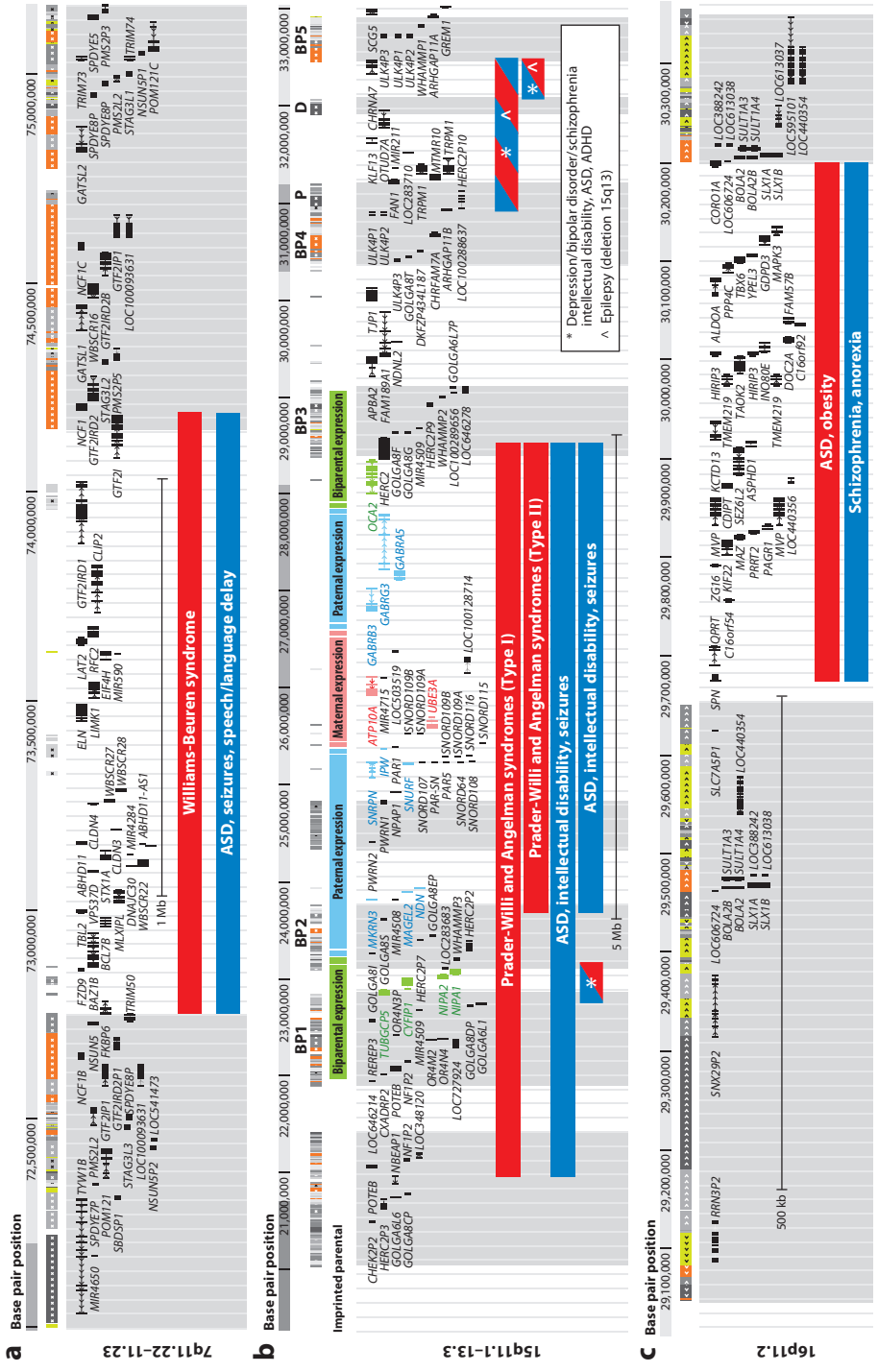


Figure 4 : Les CNVs récurrent qui sont hérité ou *de novo* chez les TSA. (a). Dans la région chromosomique 7q11, les délétions (rouge) sont associées au syndrome de Williams-Beuren, et les

duplications (bleu) sont associées aux TSA, des convulsions, et les trouble de la parole et retard de langage. (b) Dans la région chromosomique 15q11, délétions et duplications sont situés à différents points de cassure (BP1-5, P et D) constitués de séquences répétées. Les délétions maternels sont associés avec le syndrome d'Angelman, les délétions paternelles sont associées au syndrome de Prader-Willi. Les duplications maternels sont associés aux TSA, et duplications la fois maternelle et paternelle sont associés à une déficience intellectuelle et des convulsions. (c) sur la région chromosomique 16p11.2, les délétions sont associés aux TSA, à une macrocéphalie et à l'obésité, et les duplications sont associés à la schizophrénie, à la microcéphalie, et à l'anorexie. Abréviations: BP1-5, les points de cassures P et D, répétitions proximale et distale, respectivement.

Variants fréquents dans les TSA

En raison de l'absence d'hérédité mendélienne classique, l'autisme a tout d'abord considéré comme un trait polygénique impliquant de nombreux variants communs, chacun avec un petit effet. Des études de liaisons non paramétrique (sans a priori sur le modèle génétique), comme les analyses des paires de germains atteints, ont donc été réalisées et suivies par des études d'association pangénomique (GWAS). Cependant, ces approches ont identifié un nombre limité de loci, et les résultats ont été rarement reproduits (Anney, Klei et al. 2012), (Szatmari, Paterson et al. 2007, Wang, Zhang et al. 2009, Weiss, Arking et al. 2009). Le manque de résultats significatifs obtenus par GWAS a donc soulevé la question du rôle de ces variants communs dans la vulnérabilité au TSA. Pour mieux comprendre l'héritabilité manquante de traits complexes chez l'homme (Manolio, Collins et al. 2009), le groupe de Visscher (Yang, Benyamin et al. 2010) a introduit une nouvelle méthode qui tient compte de tous les SNP génotypés simultanément pour estimer la proportion de la variance pour un caractère expliqué par ces variants communs. Pour des traits quantitatifs comme la taille ou le QI, cette méthode a révélé qu'une grande partie de l'héritabilité est capturée par les SNPs, même si pour un SNP pris individuellement seulement quelques signaux significatifs ont été obtenus par GWAS (Yang, Benyamin et al. 2010, Davies, Tenesa et al. 2011). Klei et al.

(Klei, Sanders et al. 2012) a récemment utilisé cette approche quantitative de la génétique pour estimer l'héritabilité de la TSA dans un large échantillon de plus de 2.000 familles ayant un TSA et 3.600 contrôles. Ils ont constaté des effets génétiques additifs importants expliquant plus de 60 % de la variance pour les personnes atteintes de TSA provenant de familles multiplex et environ 40% pour ceux issus de familles simplex. Cependant, une étude plus récente montre que la variance expliquée par les SNPs pourrait être plus faible (15%) (Cross-Disorder Group of the Psychiatric Genomics, Lee et al. 2013) .

L'avantage de cette méthode est qu'elle montre que les SNP peuvent effectivement capturer une fraction relativement importante de l'héritabilité des TSA. L'inconvénient est qu'elle ne fournit pas d'informations sur les gènes impliqués. D'autres méthodes ont été utilisées pour agréger les variantes d'intérêt (par exemple, les gènes d'une voie biologique donnée) et de générer un score prédictif pour le diagnostic des TSA (Skafidas, Testa et al. 2012), (Yaspan, Bush et al. 2011). Les premiers résultats obtenus par ces méthodes sont encourageants, car ils ont prédit correctement diagnostic de TSA dans 56-85 % des cas (Skafidas, Testa et al. 2012). Cependant, comme pour toutes études d'association, ces premiers résultats positifs doivent être confirmés dans les grandes cohortes d'individus avant que ces « scores prédictifs » soient utilisés dans les tests de diagnostic.

En résumé, les variants communs tels que les SNPs jouent un rôle dans la vulnérabilité aux TSA, mais individuellement ils ont des effets très faibles et sont largement distribués à travers le génome. La pertinence des «scores prédictifs " basés sur les SNP fréquents pour l'aide au diagnostic doit être confirmée sur des grands échantillons. À l'avenir, il est très probable qu'une meilleure connaissance du profil génétique de chaque individu permettra une meilleure prise en charge clinique.

Plusieurs "hits" dans les TSA

Étant donné que les mutations *de novo* ne concernent qu'une fraction des personnes atteintes de TSA (<15%), il est attendu que les variants hérités vont jouer un rôle majeur dans la vulnérabilité génétique aux TSA. En outre, des variants à risque ou au contraire protecteurs pourraient contribuer à la pénétrance incomplète et l'expressivité variable que l'on observe très chez les individus porteurs de mutations *de novo*. Plusieurs études ont démontré la présence de plus d'une mutation délétère (« multiple hits ») chez les personnes atteintes de TSA (Girirajan, Rosenfeld et al. 2010, Girirajan, Rosenfeld et al. 2012, Leblond, Heinrich et al. 2012). Dans une étude à grande échelle de 2,312 enfants connus pour porter un CNV associé à un trouble neurodéveloppemental, 10% étaient aussi porteurs d'un deuxième CNV de grande taille (Girirajan, Rosenfeld et al. 2012). Les enfants qui avaient deux CNVs considérés comme délétères étaient huit fois plus susceptibles que les témoins d'avoir un retard de développement. Aucun biais parental n'a été observé pour le premier CNV *de novo* ou hérité, mais 72 % des second CNVs étaient hérités de la mère (Girirajan, Rosenfeld et al. 2012). D'autres études ont soutenu un modèle « multiples hits ». Chez 42 porteurs d'une microdélétion 16p11.2, 10 avaient un CNV supplémentaire, une proportion significativement plus élevée par rapport aux témoins conditionnés pour un grand premier hit (10 cas sur 42 , 21 sur 471 contrôles , p = 0,000057 , odds ratio = 6,6) (Girirajan, Rosenfeld et al. 2010) . Les caractéristiques cliniques des personnes atteintes de deux CNVs étaient distincts et / ou plus sévères que celles des individus porteurs d'un seul CNV. Une autre étude a montré que trois individus atteints de TSA et porteurs d'une délétion *SHANK2 de novo* étaient également porteurs d'un autre CNV au niveau du chromosome 15q11 (Leblond, Heinrich et al. 2012). Deux portaient deux CNVs entre BP4 et BP5, dont l'un duplique le gène *CHRNA7*

et l'autre supprime le gène *ARHGAP11B.* Le troisième patient portait une délétion entre BP1 et BP2 qui supprime les gènes *CYFIP1*, *NIPA1*, *NIPA2* et *TUBGCP5*. Enfin, Sato et al. (Sato, Lionel et al. 2012) ont identifié dans une famille une délétion *SHANK1* sur le chromosome 19 chez trois hommes autistes sans déficience intellectuelle. Par contre, la même délétion *SHANK1* ne déclenchait pas d'autisme chez les femmes. Une mutation stop du gène de la gamma protocadherin du gène 11- (*PCDHGA11*) a été identifié chez tous les patients, ce qui suggère que, comme pour les personnes porteuses de mutations *SHANK2*, des mutations supplémentaires augmentent le risque de TSA. En raison du petit échantillon, il n'est pas possible de confirmer si l'interaction génétique entre *SHANK1, SHANK2* et d'autres gènes synaptiques est nécessaire pour développer un TSA. Néanmoins, ces résultats devraient inciter les chercheurs à identifier des gènes modificateurs putatifs même chez les individus porteurs de mutations *de novo*. D'ailleurs très récemment Chilian et al. (2013) ont rapporté un nouveau patient avec une délétion *SHANK2* qui portait aussi une duplication *CHRNA7* et une délétion *ARHGAP11B*. (Chilian, Abdollahpour et al. 2013)

Comme dans d'autres troubles génétiques complexes, le rôle causal des variantes héritées est difficile de déterminer. Fait intéressant, les études qui combinent la génétique et les analyses fonctionnelles ont montré à plusieurs reprises un enrichissement de variantes héritées ayant un rôle délétère chez les individus atteints de TSA. Par exemple, les variants non synonymes hérités *NLGN3, NLGN4X, SHANK2*, *SHANK3* et *NRXN1* sont associés *in vitro* à une diminution de la densité synaptique (Chih, Afridi et al. 2004, Durand, Betancur et al. 2007, Berkel, Tang et al. 2011, Durand, Perroy et al. 2011, Arons, Thynne et al. 2012, Camacho-Garcia, Planelles et al. 2012, Leblond, Heinrich et al. 2012). Plusieurs cas récessifs de TSA ont été rapportés (Chahrour, Yu et al. 2012), mais surtout chez les personnes souffrant de troubles métaboliques (Coleman

and Gillberg 2012) ou dans les familles consanguines (Morrow, Yoo et al. 2008), (Yu, Chahrour et al. 2013). Dans une cohorte de 933 cas et 869 témoins, Lim et al. (2013) a récemment identifié une augmentation du double de mutations délétères affectant les deux copies de gènes autosomiques qui affichent normalement un faible taux de perte de fonction de la variation (≤ fréquence de 5%) (Lim, Raychaudhuri et al. 2013). Sur le chromosome X, soit une augmentation de 1,5 fois semblable dans KO hémizygotes rares a été observée chez les hommes atteints de TSA. Pris dans leur ensemble, les mutations autosomiques récessives pourraient contribuer à 3% des TSA. Chez les hommes, les mutations liées à l'X pourraient rendre compte de 2% des TSA. Pour identifier les régions génomiques candidats contenant des variations récessives, Casey et al. (Casey, Magalhaes et al. 2011) ont utilisé la cartographie des haplotypes homozygote chez 1,402 trios génotypés pour 1 million de SNPs. Ils ont déterminé que les régions du haplotypes homozygotes sont enrichies dans les régions précédemment rapportées pour les TSA, ce qui suggère la présence de mutations récessives non identifiés. Une autre étude a identifié à la fois une délétion et sur l'autre allèle mutation non synonyme *DIAPH3* chacun hérité d'un parent différent. La mutation paternelle P614T du gène *DIAPH3* affecte le nombre de filopodia cellulaire *in vitro* (Vorstman, Staal et al. 2006). Des études sont actuellement en cours pour déterminer si délétions pourraient démasquer des mutations sur l'autre allèle chez des individus atteints de TSA. L'ensemble de ces résultats indiquent qu'il existe des cas récessifs des TSA et que d'autres vont certainement être signalés à l'avenir. Cependant, il reste frappant de constater que l'analyse des CNVs n'a pas pu détecter un plus grand nombre d'individus porteurs de mutations récessives (par exemple, deux CNV hérités de chacun des parents), suggérant que ce mode de transmission n'est pas prépondérant dans les TSA. Des mutations affectant les gènes neuronaux sur les deux allèles pourraient être associées à des

syndromes plus graves que TSA, comme on l'observe dans les mutations homozygotes de *CNTNAP2* ou *NRXN1* qui causent un syndrome de Pitt- Hopkins- comme autosomique récessive caractérisée par un déficit intellectuel sévère, l'épilepsie et des anomalies respiratoires (Zweier, de Jong et al. 2009).

En résumé, la présence de plusieurs mutations pourrait être la règle plutôt que l'exception dans les TSA. Le défi consiste maintenant à identifier la combinaison des gènes de vulnérabilité (que ce soit en addition ou en épistasie) qui provoquent la maladie. L'identification des allèles de protection pourrait également faire la lumière sur de cibles pertinentes pour de nouveaux traitements.

Les CNVs récurrents *de novo* ou hérités dans les TSA.

Parmi les CNVs récurrents, identifiés chez les individus atteints de TSA, beaucoup sont associés à des syndromes génétiques, telles que le syndrome de Smith-Magenis, syndrome Williams Beuren, le syndrome de Sotos, les syndromes de Prader-Willi et d'Angelman, le syndrome 15q13.3 microdélétion/duplication et le syndrome de la délétion 17q21.31 (Vorstman, Staal et al. 2006). La majorité de ces loci sont encadré de séquences dupliquées du génome, qui augmente considérablement le risque de remaniement de type CNVs. Ces CNVs peuvent être soit *de novo* ou hérité de parents non atteints. Trois loci dans les régions chromosomiques du 7q11, 15q11.2-13.3 et 16p11.2 sont fortement associé aux TSA (Ballif, Hornor et al. 2007, Kumar, KaraMohamed et al. 2008, Weiss, Shen et al. 2008, Szafranski, Schaaf et al. 2010, Sanders, Ercan-Sencicek et al. 2011). Pour tous ces moci, des CNVs hérités on été observés au moins une fois (Girirajan, Rosenfeld et al. 2012). Une des particularités de ces locus est qu'ils sont associés à des symptômes cliniques très contrastés selon que la région est supprimée ou dupliqué.

Dans la région chromosomique 7q11.23 (figure 4a), les délétion sont associées au syndrome de Williams-Beuren, un trouble neurologique rare qui touche 1 sur 10.000 nouveau-nés, caractérisé par une dysmorphie faciale, une petite taille, une déficience intellectuelle, et un profil cognitif spécifique avec des aptitudes développées dans la mémoire verbale à court terme et le langage, mais avec des difficultés dans les tâches visuo-spatiales. La même région chromosomique, lorsqu'elle est dupliqué, est associée aux TSA et au TDAH , à la déficience intellectuelle, au trouble du langage, et à l'épilepsie (Somerville, Mervis et al. 2005, Depienne, Heron et al. 2007, Sanders, Ercan-Sencicek et al. 2011). Les délétions sont pour la plupart *de novo*, mais deux tiers des duplications sont héritées (Van der Aa, Rooms et al. 2009). La région critique est de 1.5-1.8Mb et contient environ 30 gènes. L'haploinsuffisance du gène « elastin-encoding » est responsable des problèmes cardio-vasculaires, mais les gènes qui causent les troubles cognitifs reste à identifier. Néanmoins plusieurs gènes exprimés dans le cerveau sont des candidats convaincants (par exemple , *TRIM50* , *FKBP6* , *FZD9* , *STX1A* , *LIMK1* et *eIF4H*). Parmi ces gènes, trois codent pour des protéines qui participent directement à la fonction synaptique (Figure 5). STX1A forme un complexe avec SNAP25 et VAMP2 (également connu sous le nom « synaptobrevin ») et est impliqué dans l'exocytose. *LIMK1* codes une protéine sérine kinase présente dans des vésicules synaptiques et qui régule la morphologie des épines dendritiques par le remodelage du squelette d'actine. Enfin, au niveau des synapses, le gène *eIF4H* est un facteur d'initiation de la traduction qui interagie avec FMRP (la protéine mutée dans le syndrome de l'X fragile) ainsi qu'avec le facteur d'élongation EIF4E et la protéine CYFIP1, tous les deux mutés chez des patients atteints de TSA (Neves-Pereira, Muller et al. 2009) (figure 5).

Dans la région chromosomique 15q11.1-13.3 (figure 4b), il y a plusieurs région avec des séquences répété en faible nombres de copies, qui augmente les risque de réarrangements génomiques en cinq points distinct de BP1 à BP5 (Szafranski, Schaaf et al. 2010). Les délétions paternels entre BP1 et BP3 (type I), BP2 et BP3 (type II) provoquent le syndrome Prader-Willi avec une prévalence de 1 sur 15,000-30,000 naissances, tandis que les délétions maternelles des même région provoquent le syndrome d'Angelman chez 1 naissance sur 12,000-20,000 (Mabb, Judson et al. 2011). Les personnes atteintes de ces syndromes peuvent également être atteint de TSA. Sur 150 personnes atteintes du syndrome de Prader-Willi et 104 personnes atteintes du syndrome d'Angelman , le taux de TSA était de 25,3% dans le premier groupe et de 1,9% dans le second (Veltman, Craig et al. 2005). Par la suite, il a été montré que les duplications maternelles de la région 15q11.1-13.3 étaient l'une des principales causes génétiques des TSA, avec environ 1% des personnes atteintes (Cook, Lindgren et al. 1997, Depienne, Moreno-De-Luca et al. 2009, Cooper, Coe et al. 2011). Plusieurs gènes de la région entre BP2 - BP3 sont soumis à empreinte parentale, ce qui explique la différence des syndrome Prader -Willi et Angelman selon la transmission paternelle ou maternelle de la délétion. Le gène *UBE3A* est responsable du syndrome d'Angelman (Kishino, Lalande et al. 1997, Matsuura, Sutcliffe et al. 1997, Sutcliffe, Jiang et al. 1997) et code pour une ubiquitine ligase qui est exprimé exclusivement par les allèles maternels dans les neurones (Yamasaki, Joh et al. 2003) (Figure 5). Les individus porteurs de délétions entre BP1 et BP3 ont généralement un phénotype plus sévère que les personnes présentant des délétions entre BP2 et BP3 (Hartley, Maclean et al. 2005, Butler, Fischer et al. 2008). Si on s'intéressent à la région BP1- BP2 (500 kb), il a quatre gènes impliqués dans le fonctionnement du cerveau : *CYFIP1, NIPA1, NIPA2* et *TUBGCP5*. *CYFIP1* code pour une protéine qui interagit avec la protéine FMRP, inhibe la traduction

synaptique (Napoli, Mercaldo et al. 2008), (Schenck, Bardoni et al. 2003). *NIPA2* a récemment été identifié comme codant pour un transporteur de magnésium, et TUBGCP5 code pour un membre dans de complexe tubuline dans le cytosquelette. Les délétions/duplications qui sont limités à BP1- BP2 sont associés à des retards dans le développement cognitif et du langage, TSA, le TDAH, la schizophrénie (Bittel, Kibiryeva et al. 2006, van der Zwaag, Staal et al. 2010). Toutefois, les variantes du nombre de copies BP1-BP2 ne sont pas entièrement pénétrante, ils sont observés chez 0,8% des personnes atteintes de TSA (0,4% délétions et des duplications 0,4%) et à 0,38% de la population générale (0,25% des délétions et des duplications 0,13 %) (Burnside, Pasion et al. 2011).

Enfin, la région BP4 - BP5 contient six gènes (*MTMR15 , MTMR10 , TRPM1 , KLF13 , OTUD7A* et *CHRNA7*) et un *microARN* (*hsamir211*). Dans cet intervalle, il y a d'autre séquences répétées proximales et distales (P et D, respectivement). Les délétions de *CHRNA7* - qui code pour un récepteur nicotiniques α7 - sont associés à un large éventail de retards de développement, déficience intellectuelle, dépression, trouble bipolaire, TSA, TDAH et l'épilepsie (International Schizophrenia Consortium 2008, Sharp, Mefford et al. 2008, Stefansson, Rujescu et al. 2008, Ben-Shachar, Lanpher et al. 2009, Dibbens, Mullen et al. 2009, Helbig, Mefford et al. 2009, Miller, Shen et al. 2009, Pagnamenta, Wing et al. 2009, Shinawi, Schaaf et al. 2009, van Bon, Mefford et al. 2009, Guilmatre, Huguet et al. 2013). Les duplications sont également associés à un risque de déficits cognitifs, mais avec une pénétrance apparemment inférieur par rapport aux délétion (Dibbens, Mullen et al. 2009, Helbig, Mefford et al. 2009, Shinawi, Schaaf et al. 2009).

Dans la région chromosomique 16p11.2 (figure 4C), plusieurs séquences répétées augmentent le risque de réarrangements génomiques. Une délétion récurrente (BP4-BP5) de 600kb contenant ~ 29 gènes est observée chez 0,5% de personnes atteint

de TSA et dans 1 cas sur 2000 dans la population générale (Ballif, Hornor et al. 2007, Battaglia, Novelli et al. 2009, Bijlsma, Gijsbers et al. 2009, Fernandez, Roberts et al. 2009, Girirajan, Rosenfeld et al. 2010), (Kumar, KaraMohamed et al. 2008, Weiss, Shen et al. 2008). Il y a un changement dans la distribution du QI pour les porteur de la délétion (QI verbal moyenne = 74 et le QI non verbal moyenne = 83), et la majorité de ces personnes nécessitent une thérapie pour le langage (Jacquemont, Reymond et al. 2011). Plus de 80% des individus porteurs de la délétion, présentent des troubles psychiatriques dont les TSA, qui est présent chez 15% des enfants porteurs de la délétion (Jacquemont, Reymond et al. 2011). De plus, la délétion est associée à l'obésité et une macrocéphalie, alors que la duplication est associée à l'anorexie et à la microcéphalie (Jacquemont, Reymond et al. 2011). L'obésité est une comorbidité majeure présente chez 50% des porteurs de délétion dés l'âge de 7 ans, mais il n'est pas en corrélation avec le QI ou tout autre trait de comportement (Jacquemont, Reymond et al. 2011). L'épilepsie est présent dans 24% des porteurs de cette mutation et s'observe indépendamment d'autres symptômes. Parmi les 29 gènes, 17 sont des gènes candidats pour les TSA (Kumar, Marshall et al. 2009, Konyukh, Delorme et al. 2011). L'analyse de la sur- ou la sous-expression des 17 gènes dans des embryons de poissons zèbres a permis d'identifier le gène *KCTD13* comme seul gène capable d'induire une microcéphalie lorsqu'il est surexprimé (se qui se rapproche de la duplication 16p11.2) et un macrocéphalie quand il est sous exprimé (semblable à une délétion 16p11.2) (Golzio, Willer et al. 2012). *KCTD13* code PDIP1, qui interagit avec PCNA27 et donc pourrait avoir un rôle dans la régulation du cycle cellulaire au cours de la neurogenèse.

En résumé, des réarrangements chromosomiques récurrents sont associés aux TSA. Le constat frappant est de constater que des délétions et duplications conduisent à des conséquences cliniques très différentes qui indiquent qu'un bon équilibre de la dose

correcte du gène est nécessaire pour le fonctionnement du cerveau humain. D'autres études seront nécessaires pour identifier les gènes responsables (ou la combinaison des gènes responsables) qui augmentent le risque de TSA chez les individus porteurs de ces réarrangements génomiques.

Les voies biologiques associées à l'autisme

Plusieurs bases de données fournissent une annotation fonctionnelle des gènes associés aux TSA. Ici, nous avons utilisé deux bases de données qui sont régulièrement mises à jour : AutismKB (http://autismkb.cbi.pku.edu.cn) et SFARI Gene (https://gene.sfari.org). Un total de 197 gènes sont inclus dans les deux bases de données, et 481 sont également inclus dans l'une ou l'autre (255 dans AutismKB et 226 dans SFARI Gene). La principale différence entre les deux bases de données concerne la sélection des gènes. AutismKB sélectionne généralement des gènes issus d'analyses de liaison, de CNVs, et de GWAS, alors que SFARI Gene sélectionne généralement des gènes issus d'études de CNVs, de séquençage de grandes cohortes et d'étude de cas. La figure 6 illustre les résultats pour 203 gènes pour lesquels des mutations *de novo* ont été identifié chez des individus atteints de TSA. Une analyse préliminaire regroupant des annotations fonctionnelles à l'aide de GeneMANIA (http://www.genemania.org) et DAVID (http://david.abcc.ncifcrf.gov) indique que 36% des protéines ont au moins une interaction (directe ou indirecte) avec une autre protéine, 61% sont exprimées dans le cerveau, et 14% sont connues pour être impliquées dans la fonction synaptique.

Les formes monogéniques de TSA comprennent des maladies génétiques telles que la sclérose tubéreuse de Bourneville (associée à *TSC1* et *TSC2*), la neurofibromatose (associée à *NF1*), le syndrome de Rett (associé à *MECP2*), et le syndrome de Cowden (associé à *PTEN*). *FMR1*, le gène muté dans le syndrome de l'X fragile, est le seul gène

dont la mutation est la plus couramment observée dans les TSA, ce qui représente environ 1-2% des cas. Des mutations rares ont également été identifié dans des gènes synaptiques, tels que *NLGN3* et *NLGN4X* ((Jamain, Quach et al. 2003)); *SHANK1*, *SHANK2*, et *SHANK3* (Durand, Betancur et al. 2007, Leblond, Heinrich et al. 2012, Sato, Lionel et al. 2012); *NRXN1* (Szatmari, Paterson et al. 2007) et *CNTNAP2* (Arking, Cutler et al. 2008, Bakkaloglu, O'Roak et al. 2008). Plus récemment, des études de mutations du gène *CHD8*, qui est impliqué dans le remodelage de la chromatine (O'Roak, Vives et al. 2012, Scheffer, Grinton et al. 2012) et du gène *TMLHE*, qui catalyse la biosynthèse de la carnitine (Celestino-Soper, Violante et al. 2012, Nava, Lamari et al. 2012), ont confirmé que les fonctions biologiques associées aux TSA vont au-delà de celles qui sont impliquées directement dans la fonction synaptique et affectent une large gamme de processus cellulaires.

Plusieurs analyses de voies biologiques ont été effectué en utilisant des données génétiques ou de transcriptome afin de mieux comprendre les fonctions biologiques associées aux TSA. Pinto et al. (2010) ont analysé la charge de CNVs chez 996 personnes atteintes de TSA et ont trouvé un enrichissement de CNVs interrompant des gènes impliqués dans la prolifération et la motilité cellulaire, ainsi que la signalisation GTPase/Ras. Iossifov et al. (2012) ont analysé les données de séquençage d'exome de 343 familles atteintes de TSA et ont observé que les mutations *de novo* étaient enrichies dans des gènes codant pour des protéines associées à la voie de la protéine FMRP (Iossifov, Ronemus et al. 2012). Voineagu et al. (2012) ont analysé les gènes qui sont exprimés de manière différentielle entre deux régions du cerveau (lobes frontaux et temporaux) chez des personnes atteintes de TSA et des témoins (Camacho-Garcia, Planelles et al. 2012). De manière intéressante, les différences régionales typiques entre les profils d'expression génique des lobes frontaux et temporaux étaient atténuées chez

les individus atteints de TSA. Un premier module du réseau lié aux interneurones et aux gènes impliqués dans la fonction synaptique était réprimé dans les cerveaux de personnes atteintes de TSA par rapport à ceux des contrôles ; un second module enrichi en gènes liés à l'immunité et à l'activation de la microglie était lui régulé à la hausse dans les cerveaux de personnes atteintes de TSA comparés à ceux des témoins. Sur la base de ces résultats génétiques et des études de modèles animaux (voir ci-dessous), trois voies se dessinent (Bourgeron 2009).

La première voie a été suggérée par les rapports de traduction aberrante d'ARNm de protéines synaptiques (Kelleher and Bear 2008). Les gènes supportant cette théorie sont *TSC1, TSC2* ,*NF1*, et *PTEN* qui codent pour des protéines qui normalement inhibent la traduction par la voie de signalisation PI3K-mTOR. Une deuxième série de gènes *FMR1, CYFIP1* et *EIF4E* sont aussi directement impliquées dans l'inhibition de la traduction des ARNm au niveau de la synapse (Figure 5).

La seconde voie concerne les gènes impliqués dans la régulation de la chromatine comme *MECP2* et *CHD8*. Cette voie est surtout associée à des formes sévère de TSA.

La troisème voie associée aux TSA concerne la balance excitation/inhibition. Plusieurs gènes associés aux TSA, comme *NLGN, NRXN* et *SHANK*, sont impliqués dans la formation des synapses excitatrices et inhibitrices. De plus, des mutations de gènes associés à l'épilepsie, tels que *SCN1A*, qui code pour un canal sodique voltage-dépendant, ont également été trouvées chez des individus atteints de TSA.

Ces trois voies (remodelage de la chromatine, traduction d'ARNm et balance excitation/inhibition) sont des cibles potentielles pour des médicaments et des essais cliniques sont en cours pour déterminer si leur ciblage pourrait améliorer les symptômes de personnes atteintes de TSA ((Jacquemont, Curie et al. 2011) (Delorme, Ey et al. 2013).

En résumé, les gènes associés aux TSA sont nombreux et impliqués dans de nombreuses fonctions cellulaires, incluant le remodelage de la chromatine, le métabolisme, la traduction de l'ARNm, et la fonction synaptique. Les conséquences en aval des mutations, cependant, pourraient converger vers un défaut dans l'homéostasie neuronale/synaptique (Ramocki and Zoghbi 2008, Toro, Konyukh et al. 2010) (Figure 5).

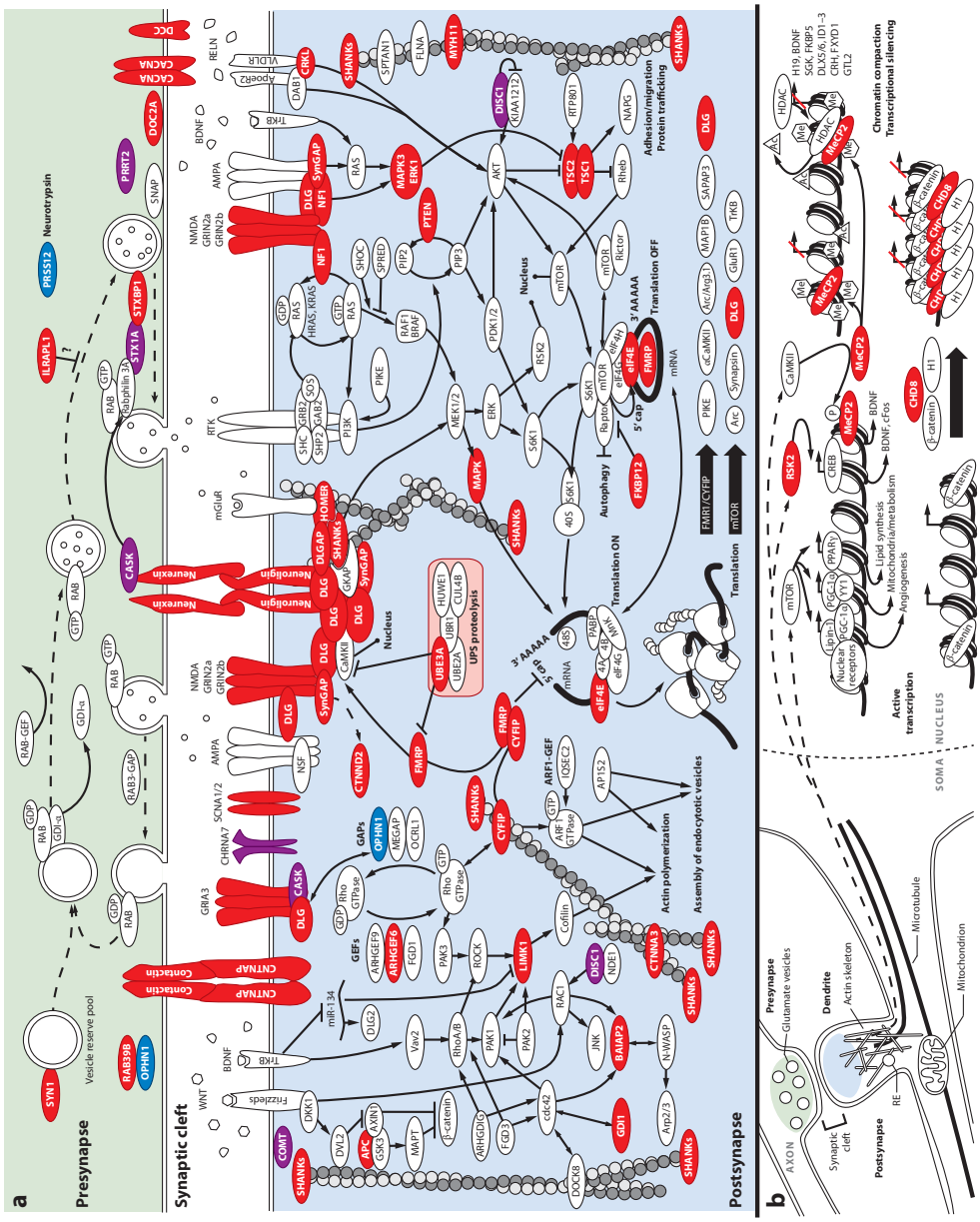


Figure 5 : Voies biologiques et protéines impliquées dans les troubles autistique. Ces protéines sont impliquées dans de nombreuses fonctions, y compris l'organisation de la densité post-synaptique, la dynamique du cytosquelette, les cascades de signalisation cellulaire, la régulation épigénétique de la transcription, et la libération de neurotransmetteurs. Les protéines associées aux TSA sont en rouge, celles associées à d'autres troubles psychiatriques sont en violet, et celles associées à la déficience intellectuelle sont en bleu. (a) Parties prés et post-synapse; (b) La synapse et la régulation de la transcription des gènes dans le noyau.

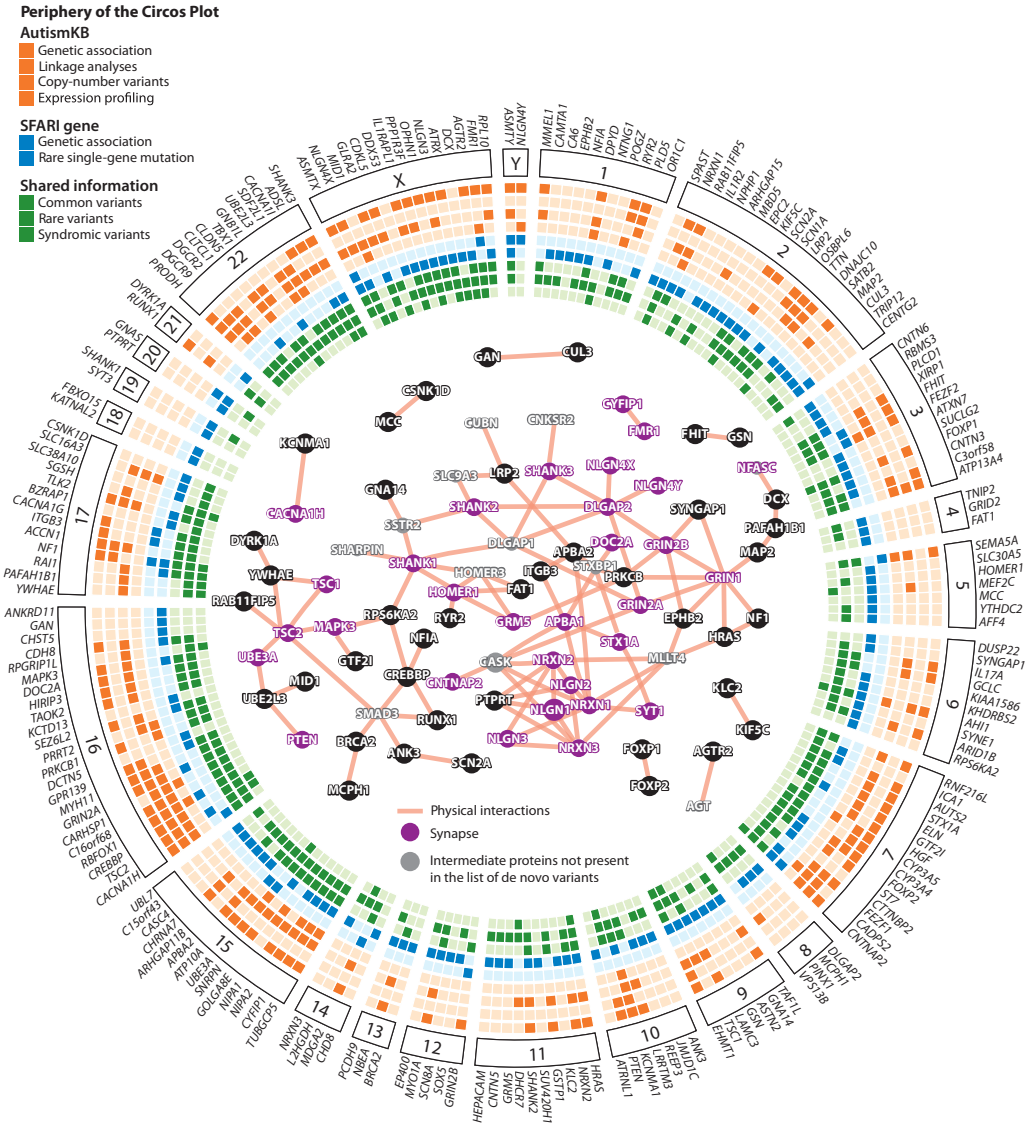


Figure 6 : Circos pots des mutations *de novo* dans les TSA. Toutes les variants codants et les CNVs présents dans AutismKB (mise à jour Juillet 2011) et SFARI Gene (mise à jour Septembre 2012) sont présentés. Une analyse de network avec GeneMANIA (au centre) met en évidence les interactions entre protéines synaptiques.

Les models souris dans l'autisme

De nombreux modèles de souris ont été développés et ont permi de caractériser les gènes mis en évidence par les études génétiques comme *PTEN , UBE3A , FMR1 , MECP2* , *mTOR, TSC* , les molécules d'adhésion cellulaire et les protéines d'échafaudage localisées spécifiquement à la synapse (Ey, Leblond et al. 2010) (figure 7). La caractérisation du comportement de ces modèles soutient aussi l'implication de ces gènes dans les principaux symptômes des TSA (Ey, Leblond et al. 2010). Les souris présentent habituellement des anomalies dans les interactions sociales (interactions libres et en cage, interactions de même sexe et de sexe opposé) et dans la communication (vocalisations ultrasoniques), ainsi que des comportements répétitifs (auto-toilettage, sauts, enterrement de billes) (Silverman, Yang et al. 2010). Fait intéressant, les modèles portant des mutations *Mecp2 , Cntnap2 , Shank2* et *Shank3* présentent des anomalies dans les trois domaines, ce qui suggère que ces gènes jouent un rôle majeur dans les circuits impliqués à la fois dans la communication sociale et les comportements répétitifs (Peca, Feliciano et al. 2011, Penagarikano, Abrahams et al. 2011, Schmeisser, Ey et al. 2012, Won, Lee et al. 2012). Les souris knockout *Fmr1 , Shank2* , présentent également une hyperactivité, une caractéristique commune chez les individus atteints de TSA (The Dutch-Belgian Fragile X Consortium 1994, Schmeisser, Ey et al. 2012, Won, Lee et al. 2012). Au niveau cellulaire, on observe à la fois des diminutions et des augmentations de la densité synaptique. Par exemple, une réduction des dendrites a été observée dans le syndrome de Rett et la sclérose tubéreuse de Bourneville, contrairement à la forte densité de dendrites relevée dans le syndrome X fragile (Auerbach, Osterweil et al. 2011). De même, certaines mutations semblaient augmenter les courants glutamate, tandis que d'autres (Schmeisser, Ey et al. 2012) les réduisent, et le même phénomène a été observé pour les courants inhibiteurs (Tabuchi, Blundell et al.

2007). De plus, plusieurs gènes associés aux TSA sont modulés par l'activité neuronale, ce qui suggère qu'ils peuvent servir d'intermédiaire dans les modifications de circuit dépendant de l'expérience (Toro, Konyukh et al. 2010). Des études récentes ont également montré des améliorations spectaculaires dans la fonction synaptique et le comportement, même chez la souris adulte (Guy, Gan et al. 2007, Baudouin, Gaudias et al. 2012). Ceci suggère, au moins dans certains cas, que le gène responsable ne joue pas un rôle central dans le développement du cerveau, mais serait plutôt nécessaire pour maintenir la fonction neurologique une fois le développement terminé. Une expérience pionnière a été réalisée par l'équipe de Bird (Guy, Gan et al. 2007), qui a créé une souris dont le gène endogène *Mecp2* a été désactivé mais pourrait potentiellement être réactivé. Grâce à cet outil, ils ont démontré que l'activation de *Mecp2* dans environ 70% des neurones à la fois chez les animaux adultes matures et immatures restauraient de nombreux défauts morphologiques dans le cortex moteur, y compris les défauts de la taille des neurones et de la complexité dendritique. Cette restauration conduit aussi à des améliorations significatives des fonctions respiratoires et sensori-motrices, y compris des modes de respiration, de la force de préhension, de la performance à la poutre et au rotarod (Guy, Gan et al. 2007) . Plus récemment, Derecki et al. (Derecki, Cronk et al. 2012) ont utilisé des greffes de moelle osseuse de type sauvage pour introduire des cellules microgliales de type sauvage dans le modèle de souris Rett et ont été en mesure d'arrêter le développement de la maladie : la durée de vie a augmenté, les modes de respiration ont été normalisés , les apnées ont été réduites , le poids corporel a presque atteint celui de type sauvage , et l'activité locomotrice a été améliorée. Enfin, les souris *NLGN3* knockout présentent une formation des synapses ectopiques et une plasticité synaptique mGluR–dépendant perturbée, une caractéristique du syndrome de l' X fragile. Ces phénotypes ont été sauvés par réexpression de *NLGN3* chez les jeunes

souris, en soulignant à nouveau la possibilité d'annuler les altérations des circuits neuronaux chez l'autiste une fois leur développement achevé (Baudouin, Gaudias et al. 2012).

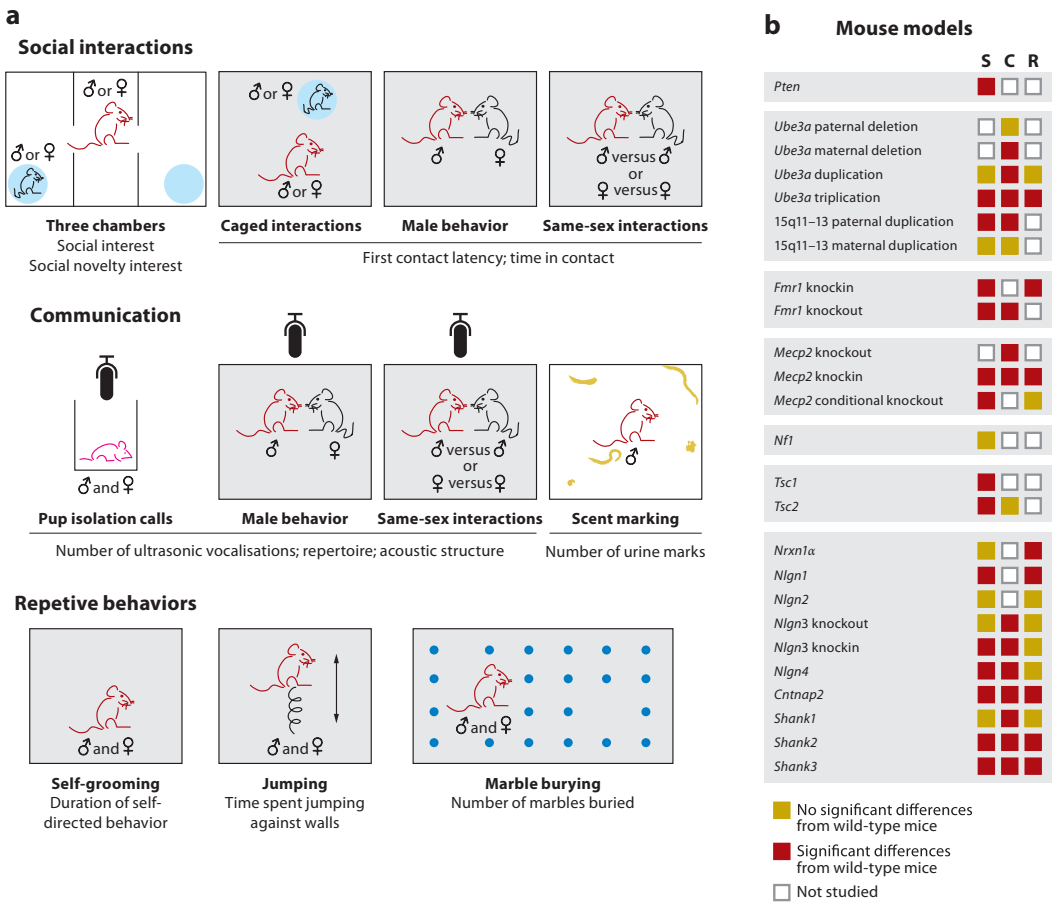


Figure 7 : Caractérisation du comportement des modèles de souris dans les TSA. (a) Les principaux tests comportementaux utilisés pour évaluer l'équivalents des trois symptômes de TSA: les interactions sociales (cage et interactions libres, des interactions de même sexe et du sexe opposé), la communication (vocalisations ultrasoniques, marquage de son territoire), et comportements répétitifs (auto-toilettage, saut, « marble burying »). (b) Bilan de la présence ou de l'absence de déficits dans les équivalents des trois symptômes de TSA de base (S, les interactions sociales, C, communication, R, comportements répétitifs) dans des modèles souris dans les TSA.

Enfin, même dans des lignées de souris génétiquement homogènes, une variabilité phénotypique peut néanmoins être observée. Par exemple, les souris knockout *NLGN4* et *NLGN3* - R451C qui présentaient des déficiences dans les interactions sociales et/ou la communication à ultrasons (Kanner 1943, Szatmari, Paterson et al. 2007) n'ont pas montré de tels déficits dans des études ultérieures (Chadman, Gong et al. 2008, Ey, Yang et al. 2012). Cette variabilité n'est pas surprenante car les individus atteints de TSA porteurs de mutations identiques présentent également différentes trajectoires cliniques. Par exemple, les individus porteurs de mutations *NLGN4X* peuvent avoir ou non une déficience intellectuelle (Jamain, Quach et al. 2003, Laumonnier, Bonnet-Brilhault et al. 2004). La prochaine étape pour mieux comprendre cette variabilité est d'étudier le rôle des facteurs épigénétiques et environnementaux, et de créer de nouveaux modèles murins porteurs d'au moins deux mutations dans des gènes différents. Ces modèles seraient plus proches de la réalité clinique des individus porteurs de mutations multiples.

En résumé, les modèles de souris porteuses de mutations dans les voies associées aux TSA peuvent présenter des anomalies dans les interactions sociales, la communication et des comportements répétitifs. La variabilité phénotypique mise en évidence dans les différents modèles n'est pas sans rappeler l'hétérogénéité clinique observée chez les individus atteints de TSA. L'idée que les phénotypes des TSA sont stables et insensibles au traitement est désormais contestée. On espère que des modèles cellulaires et animaux pourront nous dire quels gènes et voies biologiques sont affectés et donc permetront des essais cliniques plus éfficace (Delorme, Ey et al. 2013).

1.5 Perspectives pour la génétique de tsa

Au cours de la dernière décennie, d'importantes découvertes ont été faites dans la génétique des TSA avec l'identification d'un grand nombre de gènes impliqués qui affectent des voies comme la traduction de l'ARNm synaptique et la formation de dendrite et synapses. Cependant, de nombreuses questions restent sans réponse. Par exemple, le rôle des variants hérités reste difficile à déterminer, même s'ils contribuent certainement à une part importante de la sensibilité génétique aux TSA. L'excès d'hommes atteints de TSA reste largement inexpliqué, les femmes semblent atténuer l'effet des mutations plus efficacement que les hommes (qui par définition sont porteurs d'unevdélétion complète d'un chromosome X) . L'un des principaux défis dans le domaine des TSA concerne les phénotypes associés à chaque gène causal ou combinaison de gènes de vulnérabilité. Les conséquences cliniques associés à des gènes causaux dépassent les limites des TSA car les gènes associés aux TSA (par exemple, *NRXN1, SHANK3*, et *CNTNAP2*) sont également associés à d'autres troubles neuropsychiatriques tels que la schizophrénie et le trouble bipolaire (Kirov, Gumus et al. 2008, Rujescu, Ingason et al. 2009, Gauthier, Champagne et al. 2010, Verhoeven, Egger et al. 2012, Vucurovic, Landais et al. 2012), (Guilmatre, Huguet et al. 2013). Cette pléiotropie soulève la question du rôle que chaque gène joue dans le développement et le fonctionnement du cerveau humain et dans quelle mesure les différentes combinaisons d'allèles à risque pourraient mieux pronostiquer les trajectoires cliniques. Le travail collaboratif impliquant les cliniciens, les généticiens et les neurobiologistes est donc nécessaire pour mieux comprendre la diversité des individus atteints de TSA et pour améliorer leur diagnostic, leurs soins et leur intégration (Guilmatre, Huguet et al. 2013).

2. Hétérogénéité clinique et biologique de la dyslexie

2.1 - La dyslexie

Les historiens estiment à plus de 60 000 ans l'apparition du langage. Il y a seulement 5000 ans en Mésopotamie une autre révolution fit son apparition : l'invention de l'écriture. Les plus vieux documents écrits proviennent des Sumériens. Ils ne soupçonnaient pas l'importance de cette nouvelle façon de communiquer. Cet événement n'a pas été un cas isolé. En effet d'autres cultures tels que les Égyptiens, les Mayas, les Phéniciens et bien d'autres encore ont également mis au point ce système, afin de garder une trace de leur savoir ainsi que de leur histoire.

Depuis de nombreuses façons de communiquer par l'écrit ont vu le jour, les plus anciennes s'attachent à représenter le monde à l'aide de symboles, pictogrammes et idéogrammes. Dans un deuxième temps, les sons ont été représentés par des symboles, avec l'apparition des lettres, la parole est alors représentée graphiquement. Au cours du temps, l'écriture s'est développée et complexifiée, jusqu'à devenir une des bases dans notre communication. Son apprentissage (par la lecture et l'écriture) s'est vu démocratisée, et c'est à partir de l'acquisition de ces savoirs qu'on a pu mettre en evidence des inégalités au sein des individus. En effet, des personnes ne souffrant d'aucun trouble cognitif ni sensoriel, peuvent montrer de réelles difficultés dans l'apprentissage de la lecture. C'est alors que le terme dyslexie vit le jour et permit de regrouper ces différentes personnes.

Actuellement la dyslexie, est définit dans les troubles des apprentissages, plus spécifiquement dans le trouble de la lecture, il est l'un des mieux explorés, grâce à de

nombreux travaux. Dès 1978 Rutter définit avec l'aide de la World Federation of Neurology (de 1968), que « la dyslexie est un trouble manifesté par une difficulté à apprendre à lire, malgré un enseignement conventionnel, une intelligence adéquate et un bon environnement socioculturel. Elle est dépendante de troubles cognitifs fondamentaux qui sont souvent d'origine constitutionnelle ».

Ainsi les l'ensemble des troubles spécifiques des apprentissages seraient des troubles développementaux définis par des difficultés graves et persistantes d'apprentissage du langage écrit, c'est à dire de la lecture (dyslexie), de l'écriture (dysorthographie), du calcul (dyscalculie), sans étiologie retrouvée comme une déficience intellectuelle, une déficit sensoriel sévère, des lésions neurologiques, des troubles psychopathologiques graves.

Selon le DSM IV, la dyslexie appartient aux troubles spécifiques des apprentissages. C'est un trouble de l'apprentissage de la lecture et de l'écriture [annexe 2a]. L'OMS caractérise la dyslexie de développement comme : « un déficit spécifique et persistant de l'acquisition de la lecture chez un enfant ou un adulte d'intelligence normale (QI normal), recevant une instruction adéquate, en l'absence de déficits sensoriels, de pathologie psychiatrique ou neurologique avérée ou de carence psychoaffective grave.

Ainsi La dyslexie appartient aux « troubles des apprentissages », qui touchent le domaine scolaire. Ceux-ci peuvent être dus à un retard global du developpement cognitif mais peuvent aussi être dus plus spécifiquement à certains processus cognitifs (Gillberg and Soderstrorn 2003). Ainsi, à l'inverse de l'autisme, la dyslexie ne concerne pas les retards globaux du développement mais des altérations spécifiques des apprentissages. Au même titre que l'autisme, le DSM-IV a classifié ces difficultés comme des troubles intervenant durant les premières et secondes phases de l'enfance ou de l'adolescence.

Cet ensemble est constitué de différentes parties, classées selon les difficultés qu'elles regroupent: troubles de la lecture, du calcul, de l'écriture et des apprentissages non spécifiés. Cette organisation est identique à celle de la CIM-10 (Classification internationale des maladies de l'OMS, 1994) pour les altérations de l'apprentissage. De plus, il décrit pour le trouble spécifique de la lecture des critères diagnostiques adaptés à des âges différents. Le DSM-V, a revu la classification, et passe la dyslexie en arrière plan en uilisant les termes de troubles spécifiques d'apprentissage (Specific Learning Disorder) qui regroupent la dyslexie ou la dyscalculie (difficulté en mathématiques). Cet ensemble est alors inclus dans les maladies neurodeveloppemental. La raison de ce regroupement est que ces déficits se produisent fréquemment ensemble.

La dyslexie est considérée comme un trouble spécifique de l'apprentissage dont les mécanismes sont neurobiologiques (Shaywitz and Shaywitz 2005). Elle est caractérisée par des difficultés dans la reconnaissance exacte et/ou fluide de mots ainsi que par une orthographe des mots (spelling) et des capacités de décodage limitées. Ces difficultés résultent typiquement d'un déficit dans la composante phonologique du langage qui est souvent inattendue par rapport aux autres capacités cognitives de l'enfant et à l'enseignement dispensé dans sa classe. Les conséquences secondaires peuvent inclure des problèmes dans la compréhension en lecture. Cela peut entraîner une expérience réduite dans la lecture qui pourrait empêcher la croissance du vocabulaire de l'enfant et ses connaissances générales (Lyon GR 2003).

2.2 Epidémiologie

La dyslexie est un trouble de la lecture et de l'orthographe, représentant une des plus courante difficultés scolaires avec une prévalence de 5-10%. Celle-ci se manifeste lors des premières étapes de l'apprentissage de la lecture et de l'écriture. Dès lors, ce syndrôme est diagnostiqué chez des enfants ayant eu un cursus d'apprentissage normal (lecture, écriture et langage) et ne révèlant aucuns troubles visuels, auditifs, intellectuels et cognitifs sévères. Le dépistage est précoce (4 ou 5 ans) et la dysorthographie est souvent annonciatrice de ce syndrôme. Le seul traitement connu est la rééducation orthophonique et graphique. Actuellement le sex ratio est très discuté, puisque différentes études montrent que dans certains cas ce sont les garçons qui sont le plus touchés mais ce n'est pas toujours le cas (Rutter, Caspi et al. 2004). Il est probable néanmoins que les filles compenserait plus les problèmes de dyslexie.

Prévalence de la dyslexie selon les langues

La langue apprise est déterminante dans la prévalence de la dyslexie. En effet, chaque dialecte a un système d'écriture différent, et si beaucoup utilisent le même alphabet, toutes les langues ne sont pas transparentes. Une langue est dite parfaitement transparente, lorsque la correspondance entre graphèmes et phonèmes est univoque c'est à dire à un phonème correspond un et un seul graphème et reciproquement. Ainsi, les prévalences de la dyslexie sont dépendantes de la multiplicités des langues et notamment de la transparence de l'orthographe de chaque langue (Paulesu, Demonet et al. 2001).

Les travaux de Lindgren et coll. (1985) mettent en perspective cette affirmation et montrent que la prévalence de la dyslexie varie entre les États-Unis et l'Italie selon les

définitions que l'on a choisi pour caractériser ce syndrôme et la langue. Ainsi les fréquences varient de 3,6-8,5 % en Italie et de 4,5-12 % aux États-Unis. Autrement dit les taux sont significativement plus élevés aux États-Unis qu'en Italie, et ce à cause de l'orthographe, puisque contrairement à l'anglais l'italien est une langue transparente (Lindgren, De Renzi et al. 1985). Le bilinguisme ou un environnement peu stimulant pour l'enfant, p euvent aussi dans certains cas alterer ou ralentir l'apprentissage du langage. (Lindgren, De Renzi et al. 1985)

2.3 Diagnostic et évaluation

Le dépistage et le diagnostique multidiscipliniare (avec de nombreux outils (normalisés) explorent le niveau de lecture, de l'orthographe mais également une évaluation psychologique) qui passent par un entretien standardisé des parents avec l'enfant mais également des enseignants. D'autres déficits pourraient expliquer les difficultés de lecture:

- Un trouble auditif et/ou de la vue
- Un trouble de l'élocution
- Un déficit intellectuel

Au début de l'enseignement de la lecture, il est difficile de dépister la dyslexie car beaucoup d'enfants présentent, lors des premiers mois de cette apprentissage, de nombreuses difficultés qui pourraîent se confondre avec les symptômes d'une dyslexie. Néanmoins ces symptômes sont considérés comme pathologiques s'ils persistent au cours du temps. La dyslexie s'accompagne souvent d'une dysorthographie c'est à dire des difficultés dans l'apprentissage de l'écriture en plus de la lecture. Même si les règles

grammaticales sont assimilées, il y a une importante difficulté ou une impossibilité à les appliquer. Ainsi, on peut distinguer dans la dysorthographie :

- Les difficultés auditives (alors que l'audition est normale) : confusions de sons, difficulté à analyser les données auditives.
- Les difficultés dans l'organisation de l'espace, du temps, et de la phrase elle-même (il ne distingue pas les fonctions différentes des mots dans la phrase). On retrouve les erreurs spécifiques de la dyslexie.
- Confusion de lettres ou de syllabes
- Inversions
- Mauvais découpage des mots
- Méconnaissance du vocabulaire
- Méconnaissance de la grammaire

Néanmoins à l'âge adulte, une partie des dyslexiques (peu sévèrement touchés) arrivent à compenser leur handicap par des stratégies d'évitement, comme la non-utilisation de mots qui leur posent problème, en les remplaçant par des synonymes ou encore grâce à un soutien orthophonique. Cette compensation et/ou rééducation rend la détection de la dyslexie plus difficile à partir de tests classiques.

Instruments d'évaluation pour la Dyslexie

La prononciation des mots, l'association de plusieurs sons désignant une personne ou un objet, débute la deuxième année de vie ; le langage, association de mots ayant valeur de signification, doit être obtenu vers 4 ans et demi. Un premier examen orthophonique à cet âge devrait mettre en évidence la déformation persistante de mots, des phrases mal construites ou l'incapacité de retenir des phrases. Or trop souvent, la dyslexie est découverte soit lors d'un bilan mettant en evidence un échec scolaire, soit à cause d'une dysorthographie.

Une personne atteinte de dyslexie fait des erreurs, soit dans l'enchaînement des graphies, soit dans la transcription graphique des phonèmes. Elle confond la lecture de certaines lettres aux formes voisines ou proches phonétiquement comme les m et n - p, b, d, q - a et o … dès lors on peut observer ce type d'erreurs, "Piton" devient "bidon", "tacos" devient "tocas". Ces confusions ne sont pas systématiques et selon les moments, l'enfant peut lire correctement ou substituer certaines lettres ce qui entraînera une lecture hachée, hésitante, et incompréhensible. L'enfant ne réussit pas à transformer les symboles écrits en phonèmes.

Des recherches effectuées ces dernières années en neuropsychologie ont permis d'identifier un certains nombres de compétences nécessaires à l'établissement du système cognitif de lecture et/ou qui facilitent l'apprentissage de la lecture. Ces compétences sont fréquemment limitées ou déficitaires chez les enfants dyslexiques. Un certains nombres d'outils créés par des orthophonistes et/ou des neuropsychologues proposent des épreuves visant à évaluer ces compétences : des épreuves méta-phonologiques permettent d'estimer le niveau de conscience phonologique, des épreuves d'empans verbale peuvent mesurer un déficit de la mémoire à court terme. Le

traitement visuel doit aussi être évalué. Pour définir le degré de sévérité de la dyslexie différentes mesures peuvent être prises en compte, le nombre d'erreurs au cours d'une lecture mais également le temps mis pour réaliser celle-ci. Les paramètres sont ensuite étalonnés par rapport à une moyenne d'enfants non atteints mais aussi par rapport à l'âge et au niveau scolaire (baterrie odelys (outils de depistage des dyslexie), test de l'alouette).

Dans le DSM-IV ou la CIM-10 les critères sont les mêmes que ceux cités précédemment par contre la principale différence entre les deux classifications concerne la co-morbidité du trouble de la lecture avec d'autres troubles des apprentissages (notamment les troubles du calcul et de l'écriture). Ainsi pour le CIM-10 le diagnostic de trouble de la lecture est prédominant par rapport à toutes les comorbidités, tandis que le DSM-IV permet de définir plusieurs diagnostics.

2.4 Les facteurs de risque et Comorbidités

Les théories explicatives de la dyslexie sont nombreuses, faisant intervenir notament un défaut de latéralité, parmi d'autres et surtout un défaut d'attention, nécessitant une pédagogie spécifique qualitativement différente. D'autres facteurs de risques sont proposés dans l'échec de l'acquisition du langage écrit:

- Un trouble mineur lésionnel du fonctionnement cérébral ;
- Un environnement socio-culturel et économique défavorable ;
- Des méthodes d'apprentissage de la lecture inadaptées ;
- Des rythmes de progressions des acquisitions non respectés ;
- Une mauvaise formation pédagogique des maîtres ;
- Des classes surchargées etc.

En générale, 40% à 60% (McArthur, Hogben et al. 2000) (Bishop, Gobright et al. 2000) des enfants dyslexiques ont des troubles psychiatriques, comme l'anxiété, la dépression (Willcutt and Pennington 2000) et le déficit d'attention (Kadesjo and Gillberg 2001) . Les manifestations psychologiques qui accompagnent souvent les dyslexiques ont des graves effets sur les personnes qui en souffrent. Certains troubles sont souvent associés à la dyslexie :

- Trouble de l'orientation spatiale ;
- Trouble de l'orientation temporelle ;
- Trouble de la motricité oculaire (non latéralisation du regard) ;
- Défaut de discrimination et d'évocation des graphies à partir des sons ;
- Trouble de la perception du rythme de la lecture ;
- Trouble du langage avec inversion des phonèmes ;
- Trouble affectif ;
- Réactions psychologiques d'agressivité ou de passivité face à l'échec scolaire.
- Dyscalculie (Gross-Tsur, Manor et al. 1996)

2.5 Prise en charge

Il est important de détecter dès la maternelle un enfant dyslexique afin qu'il bénéficie une rééducation personnalisée avec l'aide de ses parents, médecins, orthophoniste, et dans les cas les plus sévères de l'orienter dans une classe spécialisée. Le traitement de la dyslexie repose sur deux stratégies principales : l'aide spécifique à la lecture et l'orthographe et la psychothérapie pour tout autre trouble psychologique coexistant.

Les principes de rééducation sont nombreux, et doivent être choisis selon les difficultés de l'enfant. Ils seront élaborés lors de rééducation orthophonique. La rééducation orthophonique va permettre à l'enfant de lire normalement seulement celle-ci ne fait souvent qu'une correction incomplète des difficultés. La dysorthographie constitue le handicap le plus lourd à long terme et se solde souvent par un redoublement scolaire. Dans cet apprentissage, le rôle de la dictée et de la répétition est fondamental. Le point capital de ces différentes procédures de rééducation, est que ces dernières s'inscrivent dans le temps et doivent être soutenues. De plus les troubles du graphisme nécessitent d'être analysés par des professionnels (psychomotricité, ergothérapie, neuropsychologie) afin d'être bien définis et d'apporter la rééducation la plus adéquate. Ces analyses permettent d'avoir des données quantitatives et qualitatives pour les troubles qui touchent la coordination, la gestuelle, la perception et/ou la production visuelle et visuo-spatiale.

Néanmoins, les conseils de l'Anaes (1997) sur l'orthophonie dans les troubles spécifiques du langage sont basés sur des recommandations sans preuve scientifique, puisqu'il y est dit que « La revue de la littérature n'a pas identifié d'étude permettant de proposer un nombre et une fréquence de séances fondés sur un niveau de preuve… Elle

n'a pas permis de recommander une méthode plus qu'une autre ». Cependant elles sont très largement utilisées et leur intérêt est majeur dans l'amélioration des troubles du langage, même si leur efficacité reste à être évalué.

Après l'arrêt de la rééducation (souvent à l'âge adulte), l'adaptation de la personne face à son handicap demeure indispensable au niveau de la vitesse de lecture et de la dysorthographie (M. 1993). Les nouvelles technologies tel que les outils informatiques s'avérent utiles. En effet, les correcteurs orthographiques, ou la dictée vocale présents dans les traitement de texte permettent aux personnes atteintes de compenser plus facilement.

2.6 Facteur genetique impliqué dans la dyslexie

Ainsi, si la cause n'est ni éducative, ni sociologique, quelle est-elle ? Plusieurs arguments pointent vers des causes biologiques. Les hypothèses neuro-développementales et fonctionnelles sont mises en avant comme des atteintes d'aires spécifiques du cerveau et/ou de gènes. Les données histologiques, les données d'imagerie cérébrale anatomique, et les données d'imagerie fonctionnelle convergent pour suggérer un dysfonctionnement primaire des aires périsylviennes gauches, impliquées dans la phonologie qui est au premier plan lors de l'apprentissage de la lecture (Sherman, Galaburda et al. 1985) (P. 2003) (Y. 1973).

Pour définir l'importance respective des facteurs environnementaux et génétiques pour un trouble donné, les chercheurs étudient des paires de jumeaux monozygotes ou dizygotes. Ceci permet de mesurer la concordance de la pathologie étudiée chez des personnes partageant 100% de leur génome, dans le cas des jumeaux monozygotes, ou environ 50%, dans le cas des jumeaux dizygotes. Ainsi dans la dyslexie,

lorsque d'un jumeau monozygote est dyslexique, la probabilité que l'autre le soit également est de 70 % par contre pour les jumeaux dizygotes elle n'est que de 45% (Plomin R. 1994) (K. 2001). Ces résultats permettent de d'estimer une héritabilité génétique de l'ordre de 50-60 %. Par contre ces études ne nous renseignent aucunement sur la nature et le nombre de gène mis en cause.

Les analyses de liaison dans des familles où il y a plusieurs personnes atteintes de dyslexie offre la possibilité d'aller plus loin dans la recherche génétique. Grâce à cette approche plusieurs loci on pu être identifiés dans le génomes sur différentes chromosomes. Le fait que de nombreux loci dans des région différentes du génomes ont été détecté indique qui il y a une hétérogénéité génétique avec plusieurs gènes impliqués dans la dyslexie. En outre l'hypothèse des variations *de novo* ne peut pas être rejetée car des cas sporadiques existent, ou un seul membre de la famille est atteint de dyslexie.

Les gènes impliqués dans la dyslexie

Les analyses de liaisons ou d'association génétiques portant soit sur la dyslexie ou sur des traits quantitatifs se rapportant à la lecture ont rapporté de nombreux loci. Un total de 9 régions sont référencées dans On line Mendelian Inheritance in Man (OMIM) sur les chromosomes 1p35 - 1p36 (DYX8) (Rabin, Wen et al. 1993) (Grigorenko, Wood et al. 2001) (Tzenova, Kaplan et al. 2004) (de Kovel, Franke et al. 2008) (Platko, Wood et al. 2008) , 2p11 – 2p16 (DYX3), (Fisher, Francks et al. 2002) (Fagerheim, Raeymaekers et al. 1999) (Francks, Fisher et al. 2002, Petryshen, Kaplan et al. 2002) (Kaminen, Hannula-Jouppi et al. 2003) (Peyrard-Janvid, Anthoni et al. 2004) (Anthoni, Zucchelli et al. 2007) 2q22.3 (Raskind, Igo et al. 2005) (Igo, Chapman et al. 2006) , 3p12 - 3q13 (DYX5) ,(Fisher, Francks et al. 2002) (Platko, Wood et al. 2008) (Nopola-Hemmi, Myllyluoma et

al. 2001), 4p15.33 - 4p15.32, (Bates, Luciano et al. 2007), 6p22 (DYX2) (Grigorenko, Wood et al. 1997) (Fisher, Francks et al. 2002) (Platko, Wood et al. 2008) (Cardon, Smith et al. 1994) (Kaplan, Gayan et al. 2002) (Deffenbacher, Kenyon et al. 2004), 6q12 - q15 (DYX4) (Bates, Luciano et al. 2007) (Petryshen, Kaplan et al. 2001) ,7q32 (Kaminen, Hannula-Jouppi et al. 2003) (Bates, Luciano et al. 2007) 11p15.5 (DYX7) , (Hsiung, Kaplan et al. 2004) 11q13.4 , (Raskind, Igo et al. 2005) (Igo, Chapman et al. 2006), 12q13.13-12q21.33, (Igo, Chapman et al. 2006), 13q12.13-13q12.3 , (Igo, Chapman et al. 2006) 13q22.1, (Fisher, Francks et al. 2002), 15q15 - 15q21 (DYX1) (Grigorenko, Wood et al. 1997) (Platko, Wood et al. 2008) (Smith, Kimberling et al. 1983) (Schulte-Korne, Grimm et al. 1998, Morris, Robinson et al. 2000) (Chapman, Igo et al. 2004) (Schumacher, Konig et al. 2008), 17p13.3 (Bates, Luciano et al. 2007), 18p11.2 (DYX6) (Fisher, Francks et al. 2002, Bates, Luciano et al. 2007, Seshadri, DeStefano et al. 2007) 18q22.2-18q22.3 (Fisher, Francks et al. 2002) et 21q21-21q22 (Fisher, Francks et al. 2002) et Xq26 - Xq28 (DYX9) (Fisher, Francks et al. 2002) (de Kovel, Hol et al. 2004) (Platko, Wood et al. 2008) (Bates, Luciano et al. 2007).

Au niveau de ces régions, des gènes candidats ont été analysés. La région DYX1 au niveau de 15q21 a été identifiée par analyse de liaison avec un LOD score > 3.15 (Grigorenko, Wood et al. 1997). Par la suite, grâce à deux autres familles atteintes de dyslexie, les travaux de Taipale et al. (2003) ont permis de définir le gène *DYX1C1* comme un des génes candidats dans cette région (Taipale, Kaminen et al. 2003). L'identification de deux translocations t (1; 15) (p13; q22) et t (2; 15) (q11q, 21) (Nopola-Hemmi, Taipale et al. 2000), a aussi suggéré que cette 15q21-q22 ségrége avec dyslexie. L'un des points de cassure de ces translocations est dans le gène *DYX1C1*. Le séquençage des exons de *DYX1C1*, chez 55 patients dyslexiques et 113 témoins a permis d'identifier une mutation stop aboutissant à une protéine tronquée où quatre acides

aminés manquent au niveau de l'extrémité C-terminale. L'implication de DYX1C1 et de ce locus ont été répliqué par d'autres études de liaisons (Marino, Giorda et al. 2004) (Morris, Robinson et al. 2000) (Schulte-Korne, Grimm et al. 1998). Une association entre *DYX1C1* et la dyslexie a aussi été reportée mais pas toujours répliquée (Marino, Giorda et al. 2004, Scerri, Fisher et al. 2004, Wigg, Couto et al. 2004, Bellini, Bravaccio et al. 2005, Cope, Hill et al. 2005, Meng, Hager et al. 2005, Brkanac, Chapman et al. 2007, Bates, Lind et al. 2010). Par exemple, le variant non synonyme rs17819126 (V91I) semble associé aux scores de lecture et d'orthographe sur des traitements lexicaux (Taipale, Kaminen et al. 2003). Par la suite, l'éléctroporation d'ARN interférant ciblant *DYX1C1* dans le cerveau de rats en cours de gestation a pu renseigner sur la role de *DYX1C1* au cours de la migration des neurones vers les différentes couches du cortex (Rosen, Bai et al. 2007). Ainsi l'équipe de A. Galaburda a pu constater chez ces rats, une migration neuronale altérée dans le cortex, avec des ectopies (agrégats de cellules gliales, avec généralement un regroupement de neurones) (Rosen, Bai et al. 2007).

Ces résultats font penser aux premières observations d'Albert Galaburda sur des cerveaux post-mortem de patients dyslexiques. Il avait en effet identifié des ectopies dans le cortex cérébral des patients suggérant un défaut de migration des cellules neuronal (Galaburda and Kemper 1979) (Galaburda, Sherman et al. 1985). Il a été suggéré que ces perturbations de la migration neuronale se produisent aux alentours des semaines 16 à 24 du développement chez le fétus. Dans les cerveaux, ces ectopies sont situées en grande partie dans les aires du langage (hémisphère gauche). En particulier, les aires frontales et pariéto-temporales semblent avoir plus particulièrement ces ectopies. Ainsi l'hypothèse est que le problème de connectivité impliqué dans la dyslexie (du moins pour une partie d'entre eux) serait du à ces problèmes de migration neuronale (Ramus 2004). Cette hypothèse est intéressante mais

elles est difficile à confirmer chez les patients vivants, puisqu'il n'est pas possible actuellement d'observer ces ectopies par imagerie cérébrale.

Sur le chromosome 6p une étude de liaison a permis de définir le locus DYX2, qui par la suite a été répliqué dans plusieurs études (Grigorenko, Wood et al. 1997, Fisher, Marlow et al. 1999, Grigorenko, Wood et al. 2000, Turic, Robinson et al. 2003). Ce locus contient 5 gènes *VMP, DCDC2, KIAA0319, TTRAP* et *THEM2*, mais deux gènes sont plus fréquemment proposés comme candidats, *DCDC2* (Meng, Smith et al. 2005, Schumacher, Anthoni et al. 2006) et *KIAA0319* (Cope, Harold et al. 2005) (Francks, Paracchini et al. 2004, Harold, Paracchini et al. 2006). Les arguments pour incriminé *DCDC2* est qu'il contient deux domaines peptidiques double-cortine, qui ont été initialement décrit dans la Lissencéphalie liée à l'X fragile (Dobyns, Truwit et al. 1999) avec les mutations du gène *DCX* impliqué dans la migration neuronale (Coquelle, Levy et al. 2006, Reiner, Coquelle et al. 2006). Une délétion de 2445 pb (répétitions en tandem) dans l'intron 2 de *DCDC2* a été associé significativement avec des performances de lecture (Meng, Smith et al. 2005). Cette délétion se trouve dans un site de liaison de facteur de transcription du cerveau (PEA3 and NF-ATp). D'autres études ont répliqué une association entre *DCDC2* et la dyslexie (Schumacher, Anthoni et al. 2006) (Brkanac, Chapman et al. 2007) (Lind, Luciano et al. 2010) (Wilcke, Weissfuss et al. 2009) mais pas dans tous les cas (Ludwig, Schumacher et al. 2008). Il semblerait que les hétérozygotes pour cette délétion aient un volume de matière grise significativement plus élevée dans les régions du cerveau impliqués dans le langage c'est à dire l'hémisphère gauche (Meda, Gelernter et al. 2008). De plus les zones significatives de cette augmentation semblent chevauchantes avec les données d'expression de *DCDC2* (Meda, Gelernter et al. 2008) (Meng, Smith et al. 2005) (Coquelle, Levy et al. 2006) (Reiner, Coquelle et al. 2006) (Burbridge, Wang et al. 2008). Enfin, l'interférence de *DCDC2* chez le rat provoque des anomalies de la migration

neuronale similaires à celle observées pour *DYX1C1* (Meng, Smith et al. 2005) (Burbridge, Wang et al. 2008).

Au sein de la même région, le gène *KIAA0319* est aussi candidat pour la dyslexie. Le rôle de *KIAA0319* reste inconnu mais il contient un domaine extracellulaire avec une modif MANSC (motif en N-terminale avec sept cystéines), plusieurs domaines fibronectine de type III (FN III) et un domaine « polycystic kidney disease » (PKD) qui est impliqué dans les processus d'adhérence cellulaire (Velayos-Baeza, Toma et al. 2007, Velayos-Baeza, Toma et al. 2008). KIAA0319 possède trois isoformes, dont deux sont secrétées car elles n'ont pas de domaine membranaire. L'interférence chez le rat a montré des anomalies de la migration neuronale (Paracchini, Thomas et al. 2006). Ce qui a été confirmé par un knock-down (embryonnaire) de *KIAA0319* chez les rats montrant un arrêt de la migration neuronale. En revanche, la surexpression au court de l'embryogénèse de *KIAA0319* ne provoque pas d'anomalies de la migration neuronale (Peschansky, Burbridge et al. 2010). L'expression de *KIAA0319* chez l'homme est spécifique au cerveau, particulièrement dans le cortex cérébral, l'hippocampe (CA3), le gyrus denté, putamen, l'amygdale et le cervelet (Paracchini, Thomas et al. 2006). De plus, un l'haplotype (définit par 7 SNPs) est associé à la dyslexie, surtout celui possédant l'allèle mineur de rs9461045 et rs2143340 qui sont tous les deux en déséquilibre de liaison (Dennis, Paracchini et al. 2009). L'utilisation d'un dosage par luciférase avec le promoteur de *KIAA0319*, montre que la région promotrice avec l'allèle mineur réduit le niveau d'expression de la luciférase en créant un site de liaison pour le facteur de transcription OCT-1. De plus, l'haplotype à risque est associé à seulement 40% d'expression de *KIAA0319* (Paracchini, Thomas et al. 2006). Au vu de ces données, l'hypothèse est que l'haplotype à risque fait baisser le niveau d'expression de *KIAA0319*

qui joue un rôle dans l'adhésion entre les neurones et les fibres gliales pendant la migration neuronale.

Fait intéressant, les effets génétiques pour *DCDC2* et *KIAA0319* ont été définit indépendamment l'un de l'autre et répliqué ; ce qui suggère que les deux gènes pourrait contribuer indépendamment à la dyslexie.

Le locus DYX5 sur le chromosome 3 a été mis en évidence dans une grande famille ayant 21 personnes dyslexiques, avec un LOD score > 3 (Nopola-Hemmi, Myllyluoma et al. 2001). Cette même région a été identifiée grâce à une analyse de carytotype FISH effectuée chez un patient avec infertilité qui était aussi dyslexique. L'étude a permis d'identifier une translocation réciproque équilibrée t (3; 8) (p12; q11) dans le gène *ROBO1*. Ainsi le gène *ROBO1* qui code un récepteur de guidage axonal a été proposé comme gène candidat pour la dyslexie (Hannula-Jouppi, Kaminen-Ahola et al. 2005). Par la suite l'analyse d'expression du gène par rapport à l'haplotype à risque, montre une diminution de l'expression de *ROBO1* chez les patients dyslexiques par rapport aux témoins. Ceci suggérait que l'haplo-insuffisance de ce gène pourrait augmenter le risque de dyslexie dans cette famille (Hannula-Jouppi, Kaminen-Ahola et al. 2005).

DYX3 locus (2p11-p16) contient de nombreux gènes *SEMA4F*, *OTX1*, *C2ORF3* et *MRPL19*. L'étude de Francks et al. 2002 (Francks, MacPhie et al. 2002), n'a pas montré d'association pour *SEMA4F* et *OTX1* qui sont pourtant de très bons gènes candidats puisque *SEMA4F* appartient à une famille de protéines jouant un rôle dans le guidage axonal et que *OTX1* est un facteur de transcription impliqué dans la morphogenèse du cerveau (Francks, Fisher et al. 2002). Les résultats pour les deux autres gènes semblent beaucoup plus concluant puisqu'ils sembleraient que *C2ORF3* et *MRPL19* soient

exprimés chez l'homme dans le cerveau fœtal et adulte. Mais surtout que leur expression corrèle entre *C2ORF3* et les expression de *DYX1C1*, *DCDC2* et *ROBO1*, tandis que *MRPL19* corrèle avec l'expression de *KIAA0319* (Anthoni, Zucchelli et al. 2007). Néanmoins ces résultats sur l'expression n'ont pas encore été répliqués par d'autres études. Une étude en 2012 (Thomas et al.), a testé le QI verbale et non verbale dans une cohorte de « Avon Longitudinal Study of Parents and Children » (ALSPAC) (N>5000). Seul le locus MRPL19/C2ORF3 montre une association statistiquement significative (P=0,00009) qui a été répliquéé dans quatre autres cohortes. De plus *MRPL19/C2ORF3* serait associés au volume de la matière blanche dans le corps calleux postérieur, qui relie une grande partie du cortex pariétal dans les lobes occipitaux et temporels (The Dyslexia Candidate Locus on 2p12 Is Associated with General Cognitive Ability and White Matter Structure Thomas S. Scerri 2012).

Le locus DYX9 (Xq26 - Xq28) est le seul locus pour la dyslexie présent sur le chromosome X. Il contient 15 gènes dont les gènes *SLITRK2* et *SLITRK4* qui sont considérés comme de très bons gènes candidats du fait de leur homologie avec les SLIT qui interagissent avec les protéines ROBO (Hannula-Jouppi, Kaminen-Ahola et al. 2005) (de Kovel, Hol et al. 2004). En 2013 une étude sur 12 familles multiplexes (58 sujets) a répliqué l'étude de liaisons de DYX9 en atteignant un LOD score multipoint maximum de 3,9 entre rs12558359 et rs454992 réduisant ainsi la région à 4Mb et ainsi le nombre de gènes candidats de cette dernière. Aucunes mutations ou polymorphismes qui co-ségrégent avec la dyslexie n'a été trouvée chez 7 gènes candidats (*CXORF1, CXORF51, SLITRK2*, *FMRP*, *FMR2*, *ASFMR1*, *FMR1NB*) (Huc-Chabrolle, Charon et al. 2013).

De façon paradoxale, *FMRP* pourrait être un bon gène candidat pour la dyslexie. Ce gène est responsable du syndrôme de l'X fragile avec une amplification aberrante (> 55 répétitions trinucléotides (CGG) dans son UTR 5') qui provoque l'hyperméthylation

de gène et son absence de transcription. De façon intéressante, les femmes avec un syndrome X fragile sans déficiences mentales montrent des troubles de la lecture et des difficulté en mathématiques (Bailey, Raspa et al. 2009) (Wolff, Gardner et al. 1988) (Semenza, Bonollo et al. 2012). les garçon porteurs d'une 'premutation' de FMR1 (56< prémutation< 200) n'ont pas de déficience intellectuelle, mais présentent des déficits de la mémoire de travail verbale spécifique (Cornish, Kogan et al. 2009). Le rôle de *FMRP* dans la régulation de la traduction au niveau de la synapse est majeur et une anomalie dans cette voie pourrait participer à la vulnérabilité à la dyslexie.

Les altérations du chromosome X dans la dyslexie ne se limitent pas à la région DYX9. En effet, on remarque une augmentation de la prévalence des aneuploïdies chromosomiques dans les troubles spécifiques du langage et de la dyslexie (Simpson, Addis et al. 2013). Ainsi il se peut que d'autre gène en plus de *FMRP* puisse jouer un rôle dans la dyslexie.

Les analyses de CNVs et de gènes candidats

Comme on a pu le voir précédemment, l'analyse de grandes familles et de remaniements du génome ont permis d'identifier certaines régions candidates pour la dyslexie et d'identifier des gènes candidats. D'autres études ont rapporté l'identification de CNV dans la région 21q22.3 pour les gènes *PCNT*, *DIP2A*, *S100B* et *PRMT2* chez des patients avec dyslexie (Poelmans, Engelen et al. 2009) mais également pour *DOCK4, IMMPL2* et *CNTNAP5* (Pagnamenta, Bacchelli et al. 2010) et *GABARAP, NEGR1, ACCN1, DCDC5* (Veerappa, Saldanha et al. 2013). Tous ces gènes sont de très bon candidats puisqu'ils sont directement ou plus indirectement impliqués dans la régulation de la croissance des neurites (*DIP2A* et *S100B*) (Rothermundt, Peters et al. 2003, Bai, Ghoshal et al. 2006, Guan, Haggarty et al. 2009), ainsi que dans la plasticité synaptique, comme *DOCK4* qui

régule la croissance dendritique et la ramification de l'hippocampe et de la migration (neuronale) (Ueda, Fujimoto et al. 2008) (Hiramoto, Negishi et al. 2006).

De plus une étude récente sur 13 sujets atteint d'un syndrome de microduplication du 7q11.23, ont également montré des troubles spécifiques du langage (Van der Aa, Rooms et al. 2009). Par la suite l'étude de l'expression des gènes *GTF2I* et *CYLN2* présent dans la duplication a montré une augmentation de leur expression (Van der Aa, Rooms et al. 2009). *GTF2I* code pour le facteur de transcription TFII, qui est fortement exprimé dans le cerveau (Enkhmandakh, Bitchevaia et al. 2004) (Bayarsaihan, Bitchevaia et al. 2003). De plus le promoteur du gène *DYX1C1* contient un site de liaison TFII dont un SNP (rs3743205) est associé avec la dyslexie (Taipale, Kaminen et al. 2003, Scerri, Fisher et al. 2004, Wigg, Couto et al. 2004, Marino, Citterio et al. 2007, Tapia-Paez, Tammimies et al. 2008, Dahdouh, Anthoni et al. 2009). Ainsi, l'hypothèse est que *GTF2I* pourrait réguler la transcription de *DYX1C1* (Tapia-Paez, Tammimies et al. 2008).

FOXP2* et *CNTNAP2

Les gènes *FOXP2* et *CNTNAP2* ont suscité un vif intérêt pour la communauté scientifique dans les troubles du langage mais aussi de la dyslexie. *FOXP2* code une protéine « forkhead box P2 » qui est un facteur de transcription putatif contenant un « polyglutamine tract » et un domaine de liaison d'ADN forkhead (Lai, Fisher et al. 2001). Vernes et al. (2008) ont montré que *FOXP2* régule négativement l'expression du gène *CNTNAP2* dans le cortex humain en développement(Vernes, Newbury et al. 2008). *CNTNAP2* code pour une protéine d'adhérence neuronale assimilée aux neurexines qui est exprimé dans le cerveau humain. De plus, il a été montré que *FOXP2* se lie à plusieurs séquences régulatrices présentes dans les introns de *CNTNAP2* (Vernes, Newbury et al. 2008). *CNTNAP2* joue un rôle dans les interactions neurones-glie mais aussi dans le

regroupement des canaux potassium sur les axones myélinisés permettant la transmission du potentiel d'action (Poliak, Gollan et al. 1999) (Poliak, Salomon et al. 2003).

Le gène *FOXP2* a été identifié grâce à un grand pédigrée (la famille KE) dont plusieurs membres présentaient une dyspraxie verbale sur 3 générations. En 1998, le gène a été localisé sur le chromosome 7 (Fisher, Vargha-Khadem et al. 1998) et la découverte d'un patient indépendant ayant des symptômes très similaires à la famille KE et porteur d'une translocation sur le chromosome 7 a permis l'identification du gène *FOXP2* (Lai, Fisher et al. 2001). Tous les membres atteints de la famille KE ont une mutation hétérozygote R553H localisée dans une partie conservée de la protéine.

Les analyses de l'évolution du gène *FOXP2* chez l'homme, montrent que deux substitutions T303N et N325S, semblent avoir subi une sélection positive (Enard et al . (2009). De plus, l'introduction de ces substitutions dans *Foxp2* chez la souris montre des différences qualitative dans les vocalisations ultrasoniques (chez les souriceaux), mais aussi des modifications au niveau du comportement exploratoire, de la concentration de la dopamine et de la structure des réseaux neuronaux (striatum), au contraire de la souris invalidée pour le gène *Foxp2* qui montrait des effets opposés. Enard et al. (2009) ont suggéré que ces modifications de séquences ont peut-être joué un rôle dans l'évolution de la parole et du langage chez les humains avec un effet positif du fait de leur sélection.

En 2003 l'étude de Kaminen et al. identifie la liaison de *FOXP2* dans les troubles du langage (Kaminen, Hannula-Jouppi et al. 2003). Les familles montre un lod score significatif pour des troubles du langage dans 7q32 où 2 gènes candidats sont présent *SPCH1* et *FOXP2*. Par contre aucun mutation dans les régions codantes de *FOXP2* n'a été mise en évidence chez des probants avec des problèmes d'articulation et de grammaire.

D'autres mutations chez l'Homme ont été rapportées, comme par exemple dans l'étude de Rice et al . (2012), qui décrit une délétion de *FOXP2* (*MDFIC* et *PPP1R3A*) de 1,57 Mb sur le chromosome 7q31 présente chez un garçon qui souffrait de graves troubles de la parole, d'une diminution de la motricité mais aussi d'une légère déficience cognitive. Cette délétion est transmise par la mère, qui a des troubles similaires et/ou légèrement atténués. D'autres cas décrits par Zilina et al. (2012) (Zilina, Reimand et al. 2012) ont été rapportés, notamment sur 2 familles indépendantes avec des délétions 7q31 dans *FOXP2* qui présentaient des troubles de la parole et du langage ainsi que d'autres déficits neurologiques. Dans la première famille avec une délétion de 8,3 Mb sur le chromosome 7q31.1 - q31.31 incluant le gène *FOXP2*, on observe le même tableau phénotypique entre les différentes familles porteuses de ces délétions. Néanmoins des traits autistiques sont présents, ainsi qu'un retard de croissance et de développement, une dysmorphie faciale, un nystagmus et une myopie. L'IRM cérébrale a montré une atrophie cérébrale légère et de légers hyper-signaux de la substance blanche. Dans la deuxième famille la délétion est de 6,5 Mb sur 7q31, qui inclut le gène *FOXP2*. Le phénotype des porteurs du CNV est similaire à ce qui a été vu précédemment avec en plus un comportement agressif. Ainsi ces données montrent que *FOXP2* joue un rôle dans les problèmes moteurs, du langage et de la parole.

Par la suite des analyses d'association de génétique et d'imagerie à résonance magnétique fonctionnelle (IRMf) on été effectué afin de trouver une association entre des SNPs du gène *FOXP2* et les capacités de langage. Une étude portant sur 94 individus de la population générale a montré des associations significatives entre les SNPs rs6980093 et rs7799109 et l'activation du cortex frontal gauche (Pinel, Fauchereau et al. 2012), mais aussi avec le SNP rs12533005, qui serait associé significativement à la dyslexie P = 0,016 (génotype GG avec un modèle). De plus ce même SNP est également

corrélé avec l'expression du gène *FOXP2* dans l'hippocampe chez l'homme (Wilcke, Ligges et al. 2012). Enfin, une étude à testé 16 SNPs couvrant FOXP2 et quatre gènes liés à la dyslexie, avec *DCDC2*, *DYX1C1*, *KIAA0319* et *TTRAP* dans une population de schizophrènes qui ont des problèmes de langage, et chez des contrôles (Jamadar, Powers et al. 2011). Ils ont aussi fait des IRM pour définir les volumes de matière grise dans leur cohorte. Sur les 16 SNPs, aucun SNP localisé dans *FOXP2* n'est associé aux troubles du langage. Par contre, 5 pourraient être associés à des différences dans le réseau cérébelleux pour *DYX1C1* et *DCDC2*/*KIAA0319* et dans le cortex préfrontal, temporal et occipital supérieurs pour *DCDC2*.

Pour *CNTNAP2*, au-delà de son rôle dans les troubles du spectre autistique ou la schizophrénie, une association avec les troubles du langage a été décrit pour les SNPs rs2710102 et rs759178 (Whitehouse, Bishop et al. 2011). Par la suite les mêmes SNPs ont été utilisé pour une étude sur l'acquisition précoce du langage dans une grande cohorte (606 garçons et 543 filles) (Whitehouse, Bishop et al. 2011). Un haplotype rs2710102-rs759178-rs17236239-rs2538976 a été retrouvé significativement associé à l'acquisition précoce du langage (haplotype TTAA, P = 0,049 ; GCAG d'haplotype , P = 0,0014). Ainsi des variants communs entre les exons 13-15 de *CNTNAP2* pourrait influencer l'acquisition précoce du langage (Whitehouse, Bishop et al. 2011). Des études supplémentaires, mais effectuées sur des échantillons très faibles (N=49 adultes de la population générale), montrent que d'autre SNPs tels que rs7794745 pourrait moduler le traitement syntaxique et sémantique (Kos, van den Brink et al. 2012). Enfin, dernièrement une duplication de 210 kb a été mise en évidence dans une famille de dyslexiques (Veerappa, Saldanha et al. 2013) et une délétion de plus de 10Mb de *CNTNAP2* a été observée, cette fois chez un individu atteint de bégaiement (VIQ (86), PIQ (96), FSIQ (90)) (Petrin, Giacheti et al. 2010). De plus l'IRM de ce patient porteur de

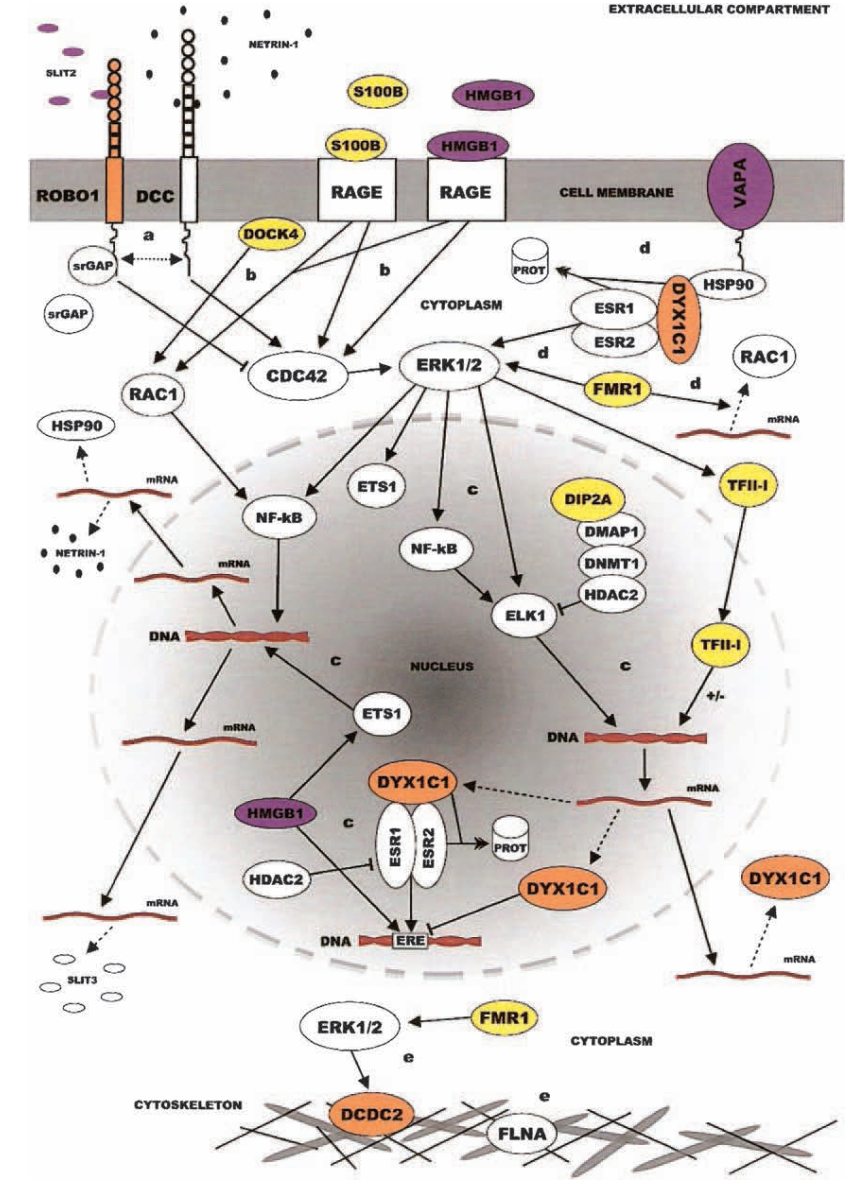
la délétion montre une atrophie cérébrale et cérébelleuse et des altérations du corps calleux.

En résumé ces deux gènes fonctionnellement liés pourraient jouer un rôle dans l'acquisition et les aptitudes linguistiques, qu'ils soient de formes syndromes rares, non syndromiques ou dans la population générale.

Les réseaux de gènes et les mécanismes candidats de la dyslexie

Comme nous l'avons décrit, de nombreuses études ont proposé des gènes candidats pour la dyslexie (*FOXP2, CNTNAP2, ROBO1, KIAA0319, S100B, DOCK4, FMR1, DIP2A, GTF2I, DYX1C1* et *DCDC2*). Les protéines codées par ces gènes pourraient être liées directement ou indirectement entre elles et aux processus de migration neuronale et/ou de croissance dirigée des neurites et axones. Dès lors l'hypothèse actuelle qui domine est que la dyslexie serait un trouble neuro-développemental causé par des anomalies dans la migration neuronale et la croissance des neurites. Cette hypothèse étayée par les modèles animaux et des donnée chez l'homme, permet de proposer plusieurs réseaux de gènes et mettre en avant certains mécanismes (Figure 8) (Poelmans, Buitelaar et al. 2011). Cependant, bien que des composants génétiques aient été mis en évidence, nous sommes loin de comprendre les mécanismes moléculaires conduisant à la dyslexie. Plusieurs gènes candidats ont été impliqués, mais il n'existe pas réellement de consensus sur un gène majeur car les résultats des réplications sont parfois mitigés. La différences entre les études et le manque de reproductibilité peut être expliqué par l'absence de mesure phénotypiques uniformes, les différentes des populations étudiées, ou encore des différences dans les analyses statistiques utilisées.

La cohorte NEURODYS à laquelle nous participons a permis de recueillir un échantillon inter-linguistique européen d'enfants qui comprend plus de 900 personnes souffrant de dyslexie. Une méta-analyse qui regroupe deux types d'analyses génétiques basées sur les SNPs et les haplotypes n'a pas pas ou peu répliqué les associations précédentes. Comme dans d'autres troubles neurocognitifs, ces résultats soulignent la nécessité de recruter de grande cohorte pour valider les effets génétiques qui sont peut-être faibles pour le cas de la dyslexie. Par conséquent, l'architecture génétique et les mécanismes neurologiques conduisant à la dyslexie restent obscurs.



Molecular Psychiatry

Figure 8: Représentation schématique du réseau de signalisation moléculaire de la dyslexie pour la croissance des neurites et la migration neuronale d'après la review de Poelmans et al. 2011 (Poelmans, Buitelaar et al. 2011). Le code de couleur utilisé dans les figures, indique le niveau de preuve qui implique les gènes / protéines candidats avec la dyslexie. En bref, les protéines orange, jaunes, sont codées par des gènes pour lesquels il existe des preuves relativement du plus au moins robuste. Les protéines pourpres

sont codées par des gènes qu'ils ont proposés comme candidats à la dyslexie par les auteurs. SLIT et Netrin regule les voies de guidage axonal des neurones par régulation de CDC42 puis par des cascades kinases (ERK 1/2) (a). La liaison de S100B et HMGB1 au récepteur RAGE entraîne l'activation de CDC42 et ERK1/2 (b). Ce dernier est également induit par FMR1. DOCK4 S100B et HMGB1 liaient à RAGE active RAC1 (b), qui à son tour active l'activité de transcription de NF-κB (c) . FMR1 active également ERK1 et ERK2 (d, e) et la traduction de l'ARNm de RAC1 (d). Activation de ERK1 et ERK2 phosphoryle NF-kB, ETS1, ELK1 et le facteur de transcription TFII-i (c). ERK1 et ERK2 également régule l'activité de DCDC2 (e). Activation de NF-kB augmente la transcription de la netrin-1 (NTN1) et HSP90, il régule également la transcription de *DYX1C1* avec *ELK1*. De plus NF-kB et ETS1 régule la transcription de SLIT3 et est directement activé par HMGB1 (c). Activée (nucléaire) TFII-I régule positivement ou négativement la transcription de *DYX1C1*, selon le contexte des autres facteurs de transcription. DIP2A se lie à DMAP1, qui est lui même liée aux complexe protéique DNMT1-HDAC2, et l'ensemble de se complexe inhibe l'activité de ELK1 sur l'expression de *DYX1C1* (via HDAC2). Ainsi HDAC2 inhibe ELK1 et ESR1 (récepteur des oestrogènes nucléaire α)(c). Lorsqu'il est lié par son ligand endogène 17β-estradiol, le ESR1 nucléaire et ESR2 (récepteur β des oestrogènes) forment un hétérodimère - augmentent l'expression de gènes contenant la séquence ERE (estrogen response elements) dans leur promoteur (c). Ce processus est amélioré, par HMGB1 et est inhibé par DYX1C1. DYX1C1 se lie aussi directement à ESR1 et ESR2, ce qui favorise leur dégradation par protéosome (c). ESR1 et ESR2 sont également présents dans le cytoplasme, où ils sont impliqués dans l'activation ERK1/2 (d). Les complexes ESR-DYX1C1 qui favorisent la dégradation protéosomale de ESR1 et ESR2 sont également présents dans le cytoplasme, et il émais l'hypothèse qu'un complexe de VAS, HSP90 et DYX1C1 serait impliqué dans la dégradation cytoplasmique ESR1 et ESR2 (d). Enfin, l'activé de DCDC2 ainsi que FLNA se lient et module directement les microtubules et l'actine dans les neurone (e). (note :HSP90, protéine de choc thermique ;NF-kB , facteur nucléaire kappa B ; VAS , VAMP associée à la protéine A.)

Résultats

Etude du gène *PRRT2* dans les troubles du spectre autistique et dans la population générale

Introduction du projet « Etude du gène PRRT2 dans les troubles du spectre autistique et la population générale» :

Les remaniements héritées et *de novo* du chromosome 16p11.2 sont associés aux troubles du spectre autistique (TSA), mais les gènes en cause restent inconnus. Parmi les gènes situés dans cette région, *PRRT2* code pour un membre du complexe synaptique SNARE qui permet la libération des vésicules. *PRRT2* est un gène candidat pour les TSA car des mutations homozygotes causent une déficience intellectuelle et les mutations hétérozygote provoquent des convulsions infantiles bénignes, des dyskinésies paroxystiques, ou migraine hémiplégique. Ici, nous avons exploré la contribution des mutations de *PRRT2* dans les TSA en criblant sa partie codante dans un large échantillon d'individus (N=1578), dont 431 personnes atteintes de TSA, 186 témoins et 961 individus issus du panel ethnique pour l'étude de la diversité du génome humain. Nous avons détecté 24 variants non synonymes, dont 1 « frameshift » (A217PfsX8) et 1 délétion de 6 pb (p.A361_P362del). La mutation du « frameshift » a été observée chez un témoin sans antécédents de troubles neurologiques ou psychiatriques. Le p.A361_P362del a été observée chez deux individus atteints d'autisme d'origine africaine sub-saharienne. Dans l'ensemble, la fréquence des variants délétères *PRRT2* n'était pas différente entre les individus atteints de TSA et les témoins. Cependant nous avons remarqué que *PRRT2* affiche un excédent très important de variants non synonymes (PN) vs synonymes (PS) en Asie (PN / PS = 4,85) et en Europe (PN / PS = 1,62) par rapport à l'Afrique (PN / PS = 0,26 ; Asie vs Afrique : P = 0.000087 ; Europe *vs* Afrique du P = 0,00035 ; Europe *vs* Asie P = P = 0,084) . Nous avons également montré qu'un traitement de l'ADN par amplification pangénomique pouvait artificiellement introduire la mutation A217PfsX8 indiquant que cette technologie ne devrait pas être effectuée avec le dépistage de mutation dans se gène *PRRT2*.

En résumé, nos résultats ne soutiennent pas un rôle pour *PRRT2* dans les TSA , mais documente sa variabilité génétique dans les populations à travers le monde. Ces informations devraient aider les chercheurs et les cliniciens à mieux étudier le rôle de PRRT2 dans les maladies.

Heterogeneous pattern of selective pressure for PRRT2 in human populations, but no association with autism spectrum disorders

Guillaume Huguet[1,2,3], Caroline Nava[4,5,6,7], Nathalie Lemière[1,2,3], Elodie Ey[1,2,3], Alexis Brice[4,5,6,8], Marion Leboyer[15,16], Pierre Szepetowski [18,19,20], Christopher Gillberg[10,11,12,17], Christel Depienne[4,5,6,8], Richard Delorme[1,2,9,14], and Thomas Bourgeron[1,2,3]*

[1] Human Genetics and Cognitive Functions, Institut Pasteur, Paris, France

[2] CNRS URA 2182 'Genes, synapses and cognition', Institut Pasteur, Paris, France

[3] University Denis Diderot Paris 7, Paris, France

[4] INSERM, U975—CRICM, Institut du cerveau et de la moelle épinière (ICM), Hôpital Pitié-Salpêtrière, Paris, France

[5] CNRS 7225—CRICM, Hôpital Pitié-Salpêtrière, Paris, France;

[6] Université Pierre et Marie Curie-Paris-6 (UPMC), UMR_S 975, Paris, France;

[7] Département de Génétique et de Cytogénétique, Unité fonctionnelle de génétique clinique, AP-HP, Hôpital Pitié-Salpêtrière, Paris, France;

[8] Département de Génétique et de Cytogénétique, Unité fonctionnelle de neurogénétique moléculaire et cellulaire, AP-HP, Hôpital Pitié-Salpêtrière, Paris, France;

[9] Assistance Publique-Hôpitaux de Paris, Robert Debré Hospital, Department of Child and Adolescent Psychiatry, Paris, France

[10] Institute of Clinical Sciences, Lund University, Malmö, Sweden

[11] Gillberg Neuropsychiatry Centre. University of Gothenburg, Göteborg, Sweden

[12] Institute of Neuroscience and Physiology, Department of Pharmacology, Gothenburg University, Sweden.

[13] Eukaryote Genotyping Platform. Genopole. Institut Pasteur, Paris, France

[14] Department of Clinical Sciences in Lund, Lund University, Lund, Sweden

[15] INSERM, U955, Psychiatry Genetic team, Creteil, France

[16] Fondation FondaMental, France

[17] Institute of Child Health, University College London, London, UK

[18] INSERM UMR_S901, Marseille, France

[19] Aix-Marseille University, Marseille, France

[20] Mediterranean Institute of Neurobiology (INMED), Marseille, France

Running title: *PRRT2* Mutations in Autism Spectrum Disorders

* Corresponding author: Thomas Bourgeron, Institut Pasteur 25 rue du docteur Roux, 75015 Paris, France; email: thomasb@pasteur.fr

Abstract:

Inherited and *de novo* genomic imbalances at chromosome 16p11.2 are associated with autism spectrum disorders (ASD), but the causative genes remain unknown. Among the genes located in this region, *PRRT2* codes for a member of the synaptic SNARE complex that allows the release of synaptic vesicles. *PRRT2* is a candidate gene for ASD since homozygote mutations are associated with intellectual disability and heterozygote mutations cause benign infantile seizures, paroxysmal dyskinesia, or hemiplegic migraine. Here, we explored the contribution of *PRRT2* mutations in ASD by screening its coding part in a large sample of 1578 individuals including 431 individuals with ASD, 186 controls and 961 individuals from the human genome Diversity Panel. We detected 24 nonsynonymous variants, 1 frameshift (A217PfsX8) and 1 in-frame deletion of 6 bp (p.A361_P362del). The frameshift mutation was observed in a control with no history of neurological or psychiatric disorders. The p.A361_P362del was observed in two individuals with autism from sub-Saharan African origin. Overall, the frequency of *PRRT2* deleterious variants was not different between individuals with ASD and controls. Remarkably, *PRRT2* displays a highly significant excess of nonsynonymous (pN) *vs* synonymous (pS) mutations in Asia (pN/pS=4.85) and Europe (pN/pS=1.62) compared with Africa (pN/pS=0.26; Asia *vs* Africa: P= 0.000087; Europe *vs* Africa P= 0.00035; Europe *vs* Asia P= P=0.084). We also showed that whole genome amplification performed through rolling cycle amplification could artificially introduce the A217PfsX8 mutation indicating that this technology should not be performed prior to *PRRT2* mutation screening. In summary, our results do not support a role for *PRRT2* coding sequence variants in

ASD, but provide an ascertainment of its genetic variability in worldwide populations that should help researchers and clinicians to better investigate the role of *PRRT2* in human diseases.

Introduction

Autism Spectrum Disorders (ASD) are characterized by impairments in reciprocal social communication, and repetitive, stereotyped and ritualistic behaviors (Association 1994). ASD include autism, Asperger syndrome and pervasive developmental disorder not otherwise specified (PDD-NOS). The prevalence of ASD overall is about 1/100, but closer to 1/300 for typical autism. The susceptibility genes to ASD remain largely unknown, but mutations affecting genes such as *NLGN3/4X*, *SHANK2/3*, *NRXN1* and *CNTNAP2* were shown to alter synaptic function and increase the risk for ASD (Huguet, Ey et al. 2013). Several large-scale studies have reported an enrichment of copy number variants (CNVs) in individuals with ASD (Pinto, Pagnamenta et al. 2010, Sanders, Ercan-Sencicek et al. 2011), suggesting that gene dosage plays a key role in the susceptibility to ASD (Toro, Konyukh et al. 2010). Among the recurrent genomic alterations associated with ASD, *de novo* or inherited submicroscopic microdeletion/duplication at 16p11.2 have been associated with a variety of developmental/neuropsychiatric disorders including ASD, intellectual disability (ID), schizophrenia, or bipolar disorders, but also with body mass index (Walters, Jacquemont et al. , Ballif, Hornor et al. 2007, Kumar, KaraMohamed et al. 2008, Weiss, Shen et al. 2008, McCarthy, Makarov et al. 2009, Barge-Schaapveld, Maas et al. 2011, Crepel, Steyaert et al. 2011, Jacquemont, Reymond et al. 2011). A meta-analysis of 6 studies including more than 2000 individuals and 30000 controls provided strong support for the role of recurrent 16p11.2 genomic imbalances as risk factors for ASD (duplication OR:20.7 and $P=1.9 \times 10^{-7}$; deletion OR:38.7 and 2.3 x

10^{-13}) (McCarthy, Makarov et al. 2009). The deleted/duplicated region at 16p11.2 spans 500-600kb and is flanked by large and highly similar 147kb low copy repeats with >99% homology, predisposing to unequal crossing-over during meiosis (Walters, Jacquemont et al. , Ballif, Hornor et al. 2007, Kumar, KaraMohamed et al. 2008, Weiss, Shen et al. 2008, McCarthy, Makarov et al. 2009). Among the 27 genes located within the interval, Kumar *et al.* (2009) screened for rare variations in 8 candidate genes selected for their function and expression in the brain (*ALDOA*, *DOC2A*, *HIRIP3*, *MAPK3*, *MAZ*, *PPP4C*, *SEZ6L2* and *TAOK2)*. None of them showed a significant association with ASD at the exception of *SEZ6L2* for which a suggestive association was detected (Kumar, Marshall et al. 2009), but not confirmed (Konyukh, Delorme et al. 2011). Smaller deletions at 16p11.2 were identified in individuals with ASD narrowing the critical region and including the *PRRT2* gene (Barge-Schaapveld, Maas et al. 2011, Crepel, Steyaert et al. 2011). *PRRT2* codes for a membrane protein that interacts with SNAP25, a protein from the synaptic vesicles release machinery (the SNARE complex). *PRRT2* is a compelling candidate gene for ASD and other psychaitric diseases since a homozygote frameshift mutation (A217PfsX8) was shown to segregate with non-syndromic ID in one consanguineous family with 9 affected individuals (Najmabadi, Hu et al. 2011). Furthermore numerous heterozygote frameshift, nonsense and missense *PRRT2* mutations (including the recurrent frameshift A217PfsX8) were identified in individuals and families with either of benign infantile seizures (BIS), paroxysmal kinesigenic dyskinesia (PKD), hemiplegic migraine (HM), or episodic ataxia, variably associated (*e.g.* the ICCA syndrome that associates BIS with PKD)(Chen, Lin et al. 2011, Cloarec, Bruneau et al. 2012, Gardiner, Bhatia et al. 2012, Lee, Huang et al. 2012). Following these results, we compared the frequency of deleterious mutations in 431 individuals with ASD and

186 controls. In addition, we ascertained the genetic variability of *PRRT2* in a sample of 961 individuals from worldwide populations.

Materials and Methods

Individuals and control samples

Mutation screening of *PRRT2* was performed in individuals with ASD recruited by the PARIS (Paris Autism Research International Sibpair) study at specialized clinical neuropsychiatric centers disposed in France and Sweden (Supplementary Table S1). Diagnosis was based on comprehensive clinical evaluation by experienced clinicians using DSM IV-TR criteria; most individuals were assessed with the Autism Diagnostic Interview-Revised (ADI-R) and some of them also with the Autism Diagnostic Observation Scale (ADOS). In Sweden, in some cases, the Diagnostic Interview for Social and Communication Disorders (DISCO-10) was used instead of the ADI-R. Cases were included only after a thorough clinical evaluation, including psychiatric and neuropsychological examination, standard karyotyping, and fragile-X testing, as well as brain imaging and EEG whenever possible. For comparison between cases and controls, only individuals from European descent were considered (Supplementary Table S2). The Human Genome Diversity Panel (HGDP) is a collection of 961 individuals from worldwide populations (Cann, de Toma et al. 2002). This study was approved by the local Institutional Review Board (IRB). The local IRB are the Comités de Protection des Personnes Île-de-France VI Sis Hôpital Pitié-Salpêtrière 75013 PARIS for France and the Sahlgrenska Academy Ethics committee, University of Gothenburg for Sweden. For all probands written inform consent was signed by the individuals or parents or the legal representative.

Mutation screening of PRRT2

We obtained DNA of 431 unrelated individuals with ASD, 180 controls and 961 individuals without known clinical status from the Human Genome Diversity Panel (HGDP) from blood leukocytes or B-lymphoblastoid cell lines, and was extracted with phenol-chloroform. All coding exons of *PRRT2* were screened by direct sequencing of the PCR products (Supplementary Table S3). The 3 coding exons of PRRT2 (NP_660282.2) and one supplementary exon present in the splicing isoforms (Q7Z6L0-1, Q7Z6L0-2, Q7Z6L0-3) were sequenced. For the amplicon of exon 2, that contains the frameshift mutation A217PfsX8, we screened an additional sample of 320 individuals with ASD and 92 controls. For all exons, the amplification of 20 ng of DNA template was performed using a standard PCR protocol with the FastStart Taq polymerase (Roche): 95°C for 15 min, followed by 35 cycles of 95°C for 30 s, 55 to 64°C (depending on the Tm) for 30 s, 72°C for 30 s to 1 min (depending on the product size), with a final cycle at 72°C for 10 min. Direct sequencing of the PCR products was performed using the BigDye Terminator Cycle V3.1 Sequencing Kit, and an ABI PRISM genetic analyzer (Applied Biosystems). For all nonsynonymous variations, the genotype was confirmed by sequencing an independent PCR product. Because of the low quantity of DNA for 3 individuals with ASD, we used whole genome amplification (WGA), using two commercially available Multiple Displacement Amplification kits. For the GenomiPhi DNA Amplification Kit (Amersham Pharmacia, Uppsala, Sweden), 25 ng of genomic DNA (in 2.5 µl) were mixed with 22.5 µl of GenomiPhi sample buffer. For the Repli-G Whole Genome Amplification kit (Qiagen Sciences Inc. Germantown, MD), 25 ng of genomic material were diluted in TE to a final volume of 2.5 µl. For comparison with the exome variant server, we verified that the coverage of the *PRRT2* exons was satisfactory (92% of the coding region had a coverage >20; Supplementary Figure 1).

Statistical analyses and software

The significance of differences in *PRRT2* variants in individuals and controls was determined by a two-sided Fisher's exact test on a two-by-two contingency table. PolyPhen-2 (http://psort.hgc.jp) and PSORT (http://genetics.bwh.harvard.edu/pph2) were used to predict the functional impact of amino acid substitutions and trans-membrane segments, respectively. To estimate the selective pressure on PRRT2 in different populations (pN/pS), the number of synonymous SNPs or nonsynonymous SNPs was divided by the total number of synonymous or nonsynonymous positions in PRRT2. The number of synonymous and nonsynonymous sites of PRRT2 were calculated by DnaSP v5 (Librado and Rozas 2009).

Results

We sequenced all coding exons of *PRRT2* in a large cohort of 1578 individuals including 431 individuals with ASD, 186 controls and 961 individuals from the HGDP (Cann, de Toma et al. 2002). Overall, we identified 24 nonsynonymous variants, 1 frameshift A217PfsX8 and 1 in-frame deletion of 6 bp (Del 361A-362P) (Table 1, Figure 1, Supplementary Table S4 and S5). Among those, 18 were not reported before (S5N, P18T, P48R, S115K, P140A, T151N, P154S, E180K, A214P, P215T, G241S, S249P, G258E, G258R, A272V, R311Q, p.A361_P362del), 1 (A217PfsX8) was repeatedly found in individuals with BIS, PDK and/or HM and 7 (E23K, P45S, P138A, D147H, P215R, P216L, R217Q) were previously observed in the general population (listed in Exome Variant Server, dbSNP or in the 1000 genomes project, Supplementary Table S6). After stratification for European ancestry, we did not observe significant enrichment of coding sequence variants in the patient sample compared with controls (Table 1). This absence

of enrichment was also observed when only variants predicted as deleterious were taken into account.

In the patient sample, we were able to ascertain the inheritance of the variants for 10 families and found that all mutations were inherited. Interestingly, the majority of the mutations were inherited from the mother (9/10; P=0.012). In two individuals with ASD from sub-Saharan African origin, a 6bp in-frame deletion (p.A361_P362del) was identified. The variant was transmitted by unaffected mothers and never observed in any individuals from our study and from other *PRRT2* mutation screening or in the 1000 genomes (>2000 controls). The genomes of these two individuals were investigated using the Illumina 1M Duo SNP array. Using Identity by state (IBS) analysis, we showed that these two individuals were not relatives (Supplementary Table 7), but were clustered with individuals from South Africa and Kenya (Supplementary Fig. 2). This *PRRT2* deletion of 2 conserved amino acids occurs in a predicted transmembrane domain of a specific isoform of *PRRT2* (Figure 1). *In silico* prediction using the PSORT software suggests however that the deletion does not disrupt the transmembrane domain of this isoform.

In the control sample, we identified one individual carrying the frameshift mutation A217PfsX8 already described as causing autosomal dominant ICCA, BIS, PKD, HM or epilepsy)(Chen, Lin et al. 2011, Cloarec, Bruneau et al. 2012, Gardiner, Bhatia et al. 2012, Lee, Huang et al. 2012). This control individual had no history of psychiatric and neurological disorders. To further explore the prevalence of A217PfsX8, we screened an additional sample of 320 individuals with ASD and 92 controls. Overall, A217PfsX8 was observed only in the control described above and not detected in any of the other 278 controls, 602 individuals with ASD and in any of the 961 individuals from the HGDP.

We used a panel of DNA from individuals originated from worldwide populations (Figure 2, Supplementary Table S4 and S5). We could identify 21 nonsynonymous *PRRT2* variants and trace the possible origin of them. For example, the predicted damaging P140A is a relatively frequent polymorphism present in Asia (7%), Oceania (5%), in Native American (7%), but more rare in Europe (allele frequency <1%). The A214P variant was observed in Asian (allele frequency <1%) and Native American populations. The P216L was observed in North Africa (allele frequency 1,7%), Europe (<1%) and Asia (<1%). Interestingly, we observed a higher ratio of nonsynonymous *vs* synonymous variants in Asia compared with Africa. In Asia, we found only 1 synonymous variant (E127E) carried by a single individual (allelic frequency of synonymous variants: 1/878=0.12%) and 15 nonsynonymous variants (S5N, P18T, P48R, P140A, D147H, T151N, P154S, A214P, P215T, P215R, P216L, G258E, A272V, R311Q) carried by 78 individuals (allelic frequency of nonsynonymous variants: 86/878=9,8%). In contrast, in Africa, we found 3 synonymous variants (P215P, L251L, C276C) carried by 7 individuals (allelic frequency of synonymous variants: 8/216=3.7%) and only 1 nonsynonymous variant (P45S) carried by 2 individuals (allelic frequency of nonsynonymous variants: 2/216=0.92%).

To evaluate selective constraints on PRRT2 in the different populations, we estimated pN/pS the ratio of nonsynonymous (pN) to synonymous polymorphism (pS) in each population sample (Table 2). Usually, most nonsynonymous changes would be expected to be eliminated by purifying selection, but under certain conditions positive or Darwinian selection may lead to their retention (Kelley and Swanson 2008). Investigating the number of synonymous and non-synonymous substitutions therefore provides information about the degree of selection operating on a gene. A pN/pS ratio greater than one implies positive or Darwinian selection; less than one implies purifying

(stabilizing) selection; and a ratio of one indicates neutral (i.e. no) selection (McDonald and Kreitman 1991). The calculation of the pN/pS ratio in the different populations revealed sign of positive or Darwinian selective pressure on PRRT2 in Asia (pN/pS=4.85) compared to Africa (pN/pS=0.17; Fischer exact test P= 0.000087). For Europe (pN/pS>1.5), there was a trend for higher nonsynonymous mutations compared to Africa (Fischer exact test P=0.057). We therefore tested if this difference was also observed between American European and American African populations available in the Exome Variant Server database (Table 2). Using this large dataset of 4300 individuals from European ancestry and 2012 individuals from African ancestry, we could confirm that there was a significant increase of nonsynonymous variants in Europe (pN/pS=1.62) compared to Africa (pN/pS=0.26) (Fischer exact test P= 0.00035). Finally, we observed a trend for higher frequency of nonsynonymous variants in Asia compared to Europe (Fischer exact test P=0.084). As expected, there was no difference between populations from the HGDP and those from EVS (Africa HGDP *vs* Africa EVS P=0.67; Europe HGDP *vs* Europe EVS P=1).

During the screening of the recurrent frameshift mutation 'A217PfsX8', we used whole genome amplified DNA for three individuals with ASD. Strikingly, we observed the A217PfsX8 mutation in these three samples for which the DNA was amplified using rolling circle amplification protocol (Genomiph from GEHealthcare) (Figure 3). We obtained the native DNA from these three individuals and found that they were actually not carrying the A217PfsX8 mutation. We then re-amplified the native DNA from these three individuals using another RCA protocol (REPLI-g Mini method from QIAGEN). After direct sequencing of the PCR products, the A217PfsX8 frameshift mutation was again present. Based on these results, we strongly advise that *PRRT2* mutation screening should not be performed on DNA previously amplified through RCA.

Discussion

Both deletions and duplications at 16p11.2 increase the risk for ASD, but the genes and the mechanisms involved remain largely unknown. Whereas several other strong candidate genes including the *KCTD13* gene are located within the deletion and can be considered possible driver(s) as well (Golzio, Willer et al. 2012), *PRRT2* was indeed a very compelling candidate. Homozygous PRRT2 mutations were associated with clinical features also observed in a subset of individuals with ASD such as intellectual disability and epilepsy. *PRRT2* encodes a putative membrane protein, localized at the synaptic membrane and interacting with SNAP25, a member of the SNARE protein complex involved in the release of synaptic vesicles. Our study was designed to identify new coding variants of *PRRT2* in a relatively large sample of individuals with ASD and to ascertain the genetic variability of the gene in worldwide populations.

We could not find any deleterious *PRRT2* mutation enrichment in individuals with ASD compared with controls. None of the patients tested were deleted or duplicated for the 16p11.2 locus. In addition, we did not find causative *PRRT2* mutations previously associated with neurological diseases such as PDK, ICCA, epilepsy or migraine in our cohort of patients. We observed a significant enrichment of maternally transmitted mutations in the individuals with ASD (9/10 transmitted by the mother). However, this apparent disequilibrium might have occurred by chance since to our knowledge, no transmission disequilibrium was reported for *PRRT2* mutations in other neurological diseases. Taken together, these results indicate that mutations in *PRRT2* are not a frequent cause of autism. Our study cannot however exclude that, in individuals with 16p11.2 deletions, haploinsufficiency of *PRRT2* could increase the risk of ASD.

Secondly, we detected the presence of the deleterious truncating mutation A217PfsX8 in a control (of note the DNA of this individual was not whole genome amplified). This man had no psychiatric diseases and, to his knowledge, no history of any psychiatric disorders in his first and second degree relatives (parents and grand parents). This individual had no diagnostic for migraine and epilepsy, but was unavailable for further clinical explorations in order to detect mild signs of neurological problems. Similar incomplete penetrance of the A217PfsX8 mutation was observed in carriers from families with ICCA syndrome (Lee, Huang et al. 2012). Based on the literature, the penetrance of the A217PfsX8 mutation was estimated to 94% (Becker, Schubert et al. 2013). It remains unknown if additional genetic variants could act as suppressors in the unaffected individuals carrying the A217PfsX8 mutation.

Thirdly, we ascertained the genetic variability of *PRRT2* in worldwide populations. We identified 21 novel nonsynonymous variants including 17 never reported before and 4 already listed in the nucleotide variant database (Supplementary Table 6). Despite the relatively large number of chromosomes tested and the population diversity, we did not observe any truncating mutations. Most of the nonsynonymous variants are restricted to one population (D147H, P154S, E180K, P215R, G241S, G258E and G258R) or to closely related populations (A214P and P216L). One variant P140A was frequently observed in different populations with a shared genetic origin (Asia, Native American and Oceania). Overall, the low rate (<1%) of nonsynonymous mutation indicates that PRRT2 is under strong selective pressure. Nevertheless, our results also revealed a heterogenous pattern of selective pressure acting on PRRT2 in human populations. The low pN/pS ratio in Africa indicates PRRT2 is under the usual purifying selection (as for the majority of the genome), whereas the pN/pS ratio in Asia and Europe is higher than expected and consistent with positive selection (Tennessen,

Bigham et al. 2012). To date, we have no explanation for this observation. It would be interesting however to see if there is a higher prevalence of PRRT2 related disease in Asia and Europe compared to Africa. It would also be interesting to explore the genetic diversity of the binding partners of PRRT2 in order to decipher whether this lower selective pressure is restricted to PRRT2 or to other members of the same pathway.

Finally, we showed that whole genome amplification using multiple displacement amplification technology such as GenomiPhi and Repli-G could lead to the appearance of the A217PfsX8 mutation. The genomic region surrounding the mutation is made of 4 guanines and 9 cytosines and seems therefore to be at risk for mutation both *in vivo* and *in vitro.* For this reason, we strongly suggest that *PRRT2* mutation screening should be restricted to native DNA and validate using Sanger sequencing. Interestingly, in the last release of the Exome Variant Server database, at the same location of the A217PfsX8 mutation (hg19, chr16: 29825015), 10% of the *PRRT2* allele contains an additional C (allele A1 in EVS) and 10% have a deletion of a C (allele A2 in EVS), but with very low average sample read depth (N=10). Given the very high penetrance of *PRRT2* in human diseases, it is impossible that 20% of the *PRRT2* alleles carry a frame shift mutation at this position. Especially, since we found the A217PfsX8 mutation in only 1 individual out of 1990 (Allelic frequency= 0.025%). This high frequency of A217PfsX8 mutation in EVS could be related to the DNA amplification step required for whole exome sequencing. Studying the mechanism leading to the A217PfsX8 mutation was beyond the scope of our study, but our observation that a DNA amplification step through RCA could introduce this additional C/G might help in understanding the mechanism leading to A217PfsX8 mutation.

In summary, our study indicates that *PRRT2* mutations do not play a major role in the susceptibility to ASD and confirm that truncating mutations of *PRRT2* are not fully

penetrant. We also provide an ascertainment of the genetic diversity of *PRRT2* in worldwide populations as well as important indication on the pitfalls for the mutation screening. All together, these results should help researchers and clinicians to better investigate the role of *PRRT2* in human diseases.

References

1. Association AP (1994) Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition. Washington, DC: APA.
2. Huguet G, Ey E, Bourgeron T (2013) The Genetic Landscapes of Autism Spectrum Disorders. Annu Rev Genomics Hum Genet 14: In press.
3. Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, et al. (2010) Functional impact of global rare copy number variation in autism spectrum disorders. Nature 466: 368-372
4. Sanders SJ, Ercan-Sencicek AG, Hus V, Luo R, Murtha MT, et al. (2011) Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. Neuron 70: 863-885.
5. Toro R, Konyukh M, Delorme R, Leblond C, Chaste P, et al. (2010) Key role for gene dosage and synaptic homeostasis in autism spectrum disorders. Trends Genet 26: 363-372.
6. Jacquemont S, Reymond A, Zufferey F, Harewood L, Walters RG, et al. (2011) Mirror extreme BMI phenotypes associated with gene dosage at the chromosome 16p11.2 locus. Nature 478: 97-102.
7. Barge-Schaapveld DQ, Maas SM, Polstra A, Knegt LC, Hennekam RC (2011) The atypical 16p11.2 deletion: a not so atypical microdeletion syndrome? Am J Med Genet A 155A: 1066-1072.
8. Crepel A, Steyaert J, De la Marche W, De Wolf V, Fryns JP, et al. (2011) Narrowing the critical deletion region for autism spectrum disorders on 16p11.2. Am J Med Genet B Neuropsychiatr Genet 156: 243-245.
9. Ballif BC, Hornor SA, Jenkins E, Madan-Khetarpal S, Surti U, et al. (2007) Discovery of a previously unrecognized microdeletion syndrome of 16p11.2-p12.2. Nat Genet 39: 1071-1073.
10. Kumar RA, KaraMohamed S, Sudi J, Conrad DF, Brune C, et al. (2008) Recurrent 16p11.2 microdeletions in autism. Hum Mol Genet 17: 628-638.
11. Weiss LA, Shen Y, Korn JM, Arking DE, Miller DT, et al. (2008) Association between microdeletion and microduplication at 16p11.2 and autism. N Engl J Med 358: 667-675.
12. McCarthy SE, Makarov V, Kirov G, Addington AM, McClellan J, et al. (2009) Microduplications of 16p11.2 are associated with schizophrenia. Nat Genet 41: 1223-1227.
13. Walters RG, Jacquemont S, Valsesia A, de Smith AJ, Martinet D, et al. A new highly penetrant form of obesity due to deletions on chromosome 16p11.2. Nature 463: 671-675.
14. Kumar RA, Marshall CR, Badner JA, Babatz TD, Mukamel Z, et al. (2009) Association and mutation analyses of 16p11.2 autism candidate genes. PLoS One 4: e4582.
15. Konyukh M, Delorme R, Chaste P, Leblond C, Lemiere N, et al. (2011) Variations of the candidate SEZ6L2 gene on Chromosome 16p11.2 in patients with autism spectrum disorders and in human populations. PLoS One 6: e17289.
16. Najmabadi H, Hu H, Garshasbi M, Zemojtel T, Abedini SS, et al. (2011) Deep sequencing reveals 50 novel genes for recessive cognitive disorders. Nature 478: 57-63.
17. Chen WJ, Lin Y, Xiong ZQ, Wei W, Ni W, et al. (2011) Exome sequencing identifies truncating mutations in PRRT2 that cause paroxysmal kinesigenic dyskinesia. Nat Genet 43: 1252-1255.

18. Cloarec R, Bruneau N, Rudolf G, Massacrier A, Salmi M, et al. (2012) PRRT2 links infantile convulsions and paroxysmal dyskinesia with migraine. Neurology 79: 2097-2103.
19. Gardiner AR, Bhatia KP, Stamelou M, Dale RC, Kurian MA, et al. (2012) PRRT2 gene mutations: from paroxysmal dyskinesia to episodic ataxia and hemiplegic migraine. Neurology 79: 2115-2121.
20. Lee HY, Huang Y, Bruneau N, Roll P, Roberson ED, et al. (2012) Mutations in the novel protein PRRT2 cause paroxysmal kinesigenic dyskinesia with infantile convulsions. Cell Rep 1: 2-12.
21. Cann HM, de Toma C, Cazes L, Legrand MF, Morel V, et al. (2002) A human genome diversity cell line panel. Science 296: 261-262.
22. Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25: 1451-1452.
23. Kelley JL, Swanson WJ (2008) Positive selection in the human genome: from genome scans to biological significance. Annu Rev Genomics Hum Genet 9: 143-160.
24. McDonald JH, Kreitman M (1991) Adaptive protein evolution at the Adh locus in Drosophila. Nature 351: 652-654.
25. Golzio C, Willer J, Talkowski ME, Oh EC, Taniguchi Y, et al. (2012) KCTD13 is a major driver of mirrored neuroanatomical phenotypes of the 16p11.2 copy number variant. Nature 485: 363-367.
26. Becker F, Schubert J, Striano P, Anttonen AK, Liukkonen E, et al. (2013) PRRT2-related disorders: further PKD and ICCA cases and review of the literature. J Neurol 260: 1234-1244.
27. Tennessen JA, Bigham AW, O'Connor TD, Fu W, Kenny EE, et al. (2012) Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science 337: 64-69.

Acknowledgments

We thank the DNA and cell bank of Cochin hospital and of the CRICM, the Plateforme de Ressource Biologique (B. Ghaleh), the Clinical Investigation Centers of Mondor-Chenevier and Robert Debré hospitals for their assistance, and the Psychiatric Genetics team (Stéphane Jamain) of Inserm U955 for collection and management of DNAs. We thank E. Patin, G. Laval and D. diMoreno for careful reading of the manuscript. We thank C. Bouchier for the use of sequencing facilities and B. Regnault and L. Lemée for the Illumina genotyping arrays at the Institut Pasteur Génopole.

Financial Disclosure

This work was funded by the Institut Pasteur, CNRS, INSERM, AP-HP, University Paris Diderot, the Bettencourt-Schueller foundation, the Orange foundation, the FondaMental foundation, the Conny-Maeva foundation, the Cognacq-Jay foundation, the ANR (ANR-08-MNPS-037-01 - SynGen), Neuron-ERANET (EUHF-AUTISM), the DFG (BO1718/3-1, 4-1) and Baustein L.SBN.0081. "The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript."

Competing Interest

The authors have declared that no competing interests exist.

Ethics statement

This study was approved by the local Institutional Review Board (IRB) and written inform consents were obtained from all participants of the study. The local IRB are the “Comité de Protection des Personnes” (Île-de-France Hôpital Pitié-Salpêtrière Paris, France); the Sahlgrenska Academy Ethics committee (University of Gothenburg, Sweden).

Figure legends

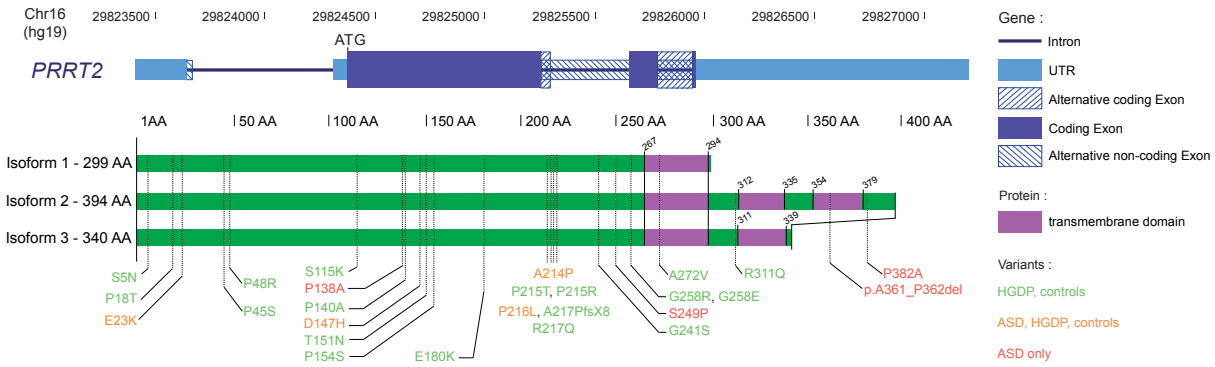


Figure 1. *PRRT2* coding variants identified in this study. Schematic diagram of the *PRRT2* gene and of the three PRRT2 protein isoforms. Mutation identified in this study are indicated in green (controls and HGDP), orange (controls, HGDP, and patients) and red (patients only).

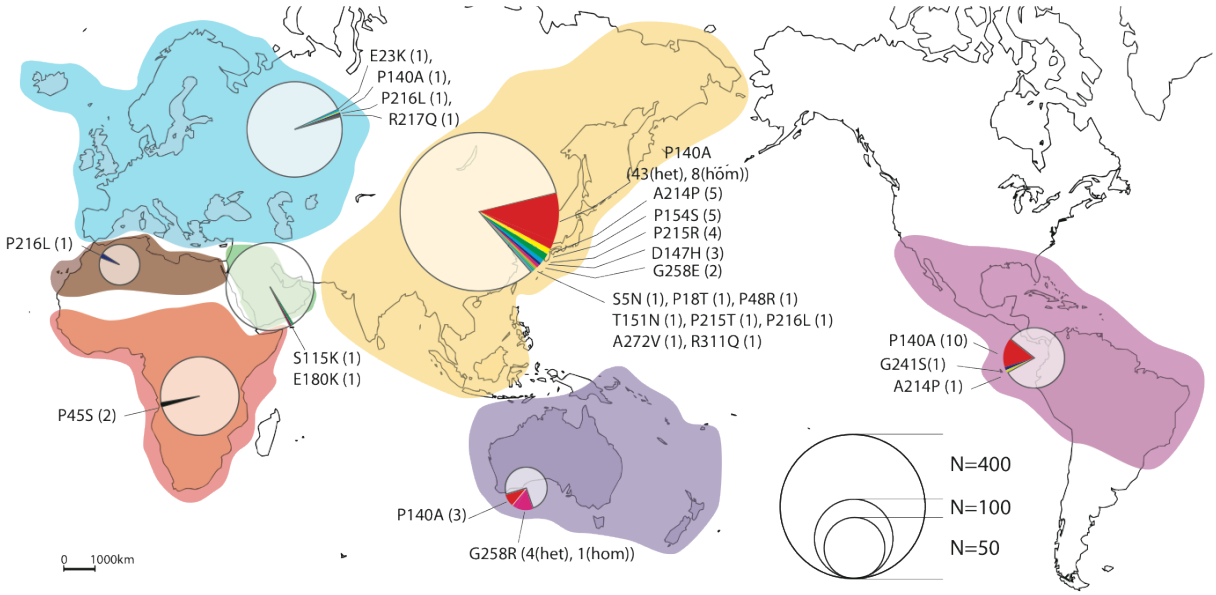


Figure 2. *PRRT2* variants identified in individuals from worldwide populations. A total of 961 individuals from the human genome diversity panel (HGDP) were sequenced for all *PRRT2* exons. The diameter of each circle is proportional to the number of individuals who were sequenced for *PRRT2*.

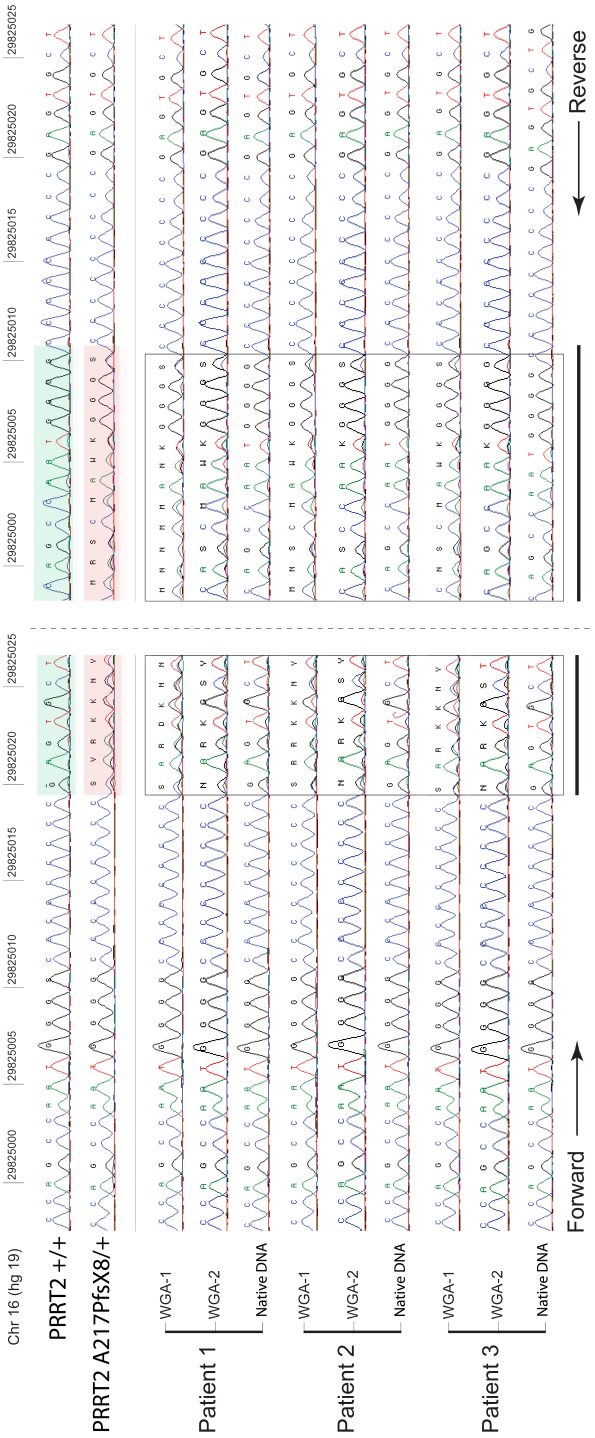


Figure 3. Chromatograms of the *PRRT2* A217PfsX8 mutation before and after whole genome amplification. The chromatograms of *PRRT2* sequence before (Native DNA) and after whole genome amplification of the DNA from three independent patients using two different protocols GenomiPhi DNA Amplification Kit (WGA-1) or Repli-G Whole Genome Amplification kit (WGA-2). The mutation was not present in the native DNA, but was detected after whole genome amplification.

Table 1. *PRRT2* coding variants identified in individuals with ASD and controls

Variants	dbSNP/EVS/1KG	Inheritance	Ethnicity	Polyphen2	ASD (All; European ancestry)	Controls
E23K[a]	1KG_16_29824442	ND	2 Europe	Benign	2/431; 2/261	0/180
P138A[a]	rs79182085	ND	1 Asia	Benign	1/431; 0/261	0/180
D147H[a]	rs79568162	Maternal	1 Asia	Benign	1/431; 0/261	0/180
A214P[a]	-	4 Maternal	1 Asia, 1 Europe, 2 Africa	Damaging	4/726; 1/474	0/272
P216L[a]	rs76335820	2 Maternal, 1 Paternal, 5 ND	6 Europe,1 Mixte, 1 ND	Damaging	8/726; 6/474	8/272
A217PfsX8	-	ND	1 Europe	Damaging	0/726; 0/474	1/272
S249P	-	ND	1 Mixte	Damaging	1/726; 0/474	0/272
p.A361_P362del	-	2 Maternal	2 Sub Saharan Africa	Damaging[b]	2/431; 0/261	0/180
P382A	-	Maternal	1 Mixte	Damaging	1/431; 0/261	0/180

[a]Variants observed in the human genome diversity panel.
[b]The p.A361_P362del was considered as probably damaging since it affects conserved amino acids in *PRRT2*.

Table 2. Synonymous and nonsynonymous variations of *PRRT2* in worldwide populations.

Ethnic	Nonsynonymous variants	Synonymous variants	pN (nonsynonymous variants / nonsynonymous sites)[a]	pS (synonymous variants / synonymous sites)[a]	pN/pS
EVS Europe (N=4300)	14	3	0.021	0.013	1.62
EVS African-American (N=2012)	6	8	0.009	0.035	0.26
HGDP Asia (N=439)	14	1	0.021	0.004	4.85
HGDP Africa (N=108)	1	2	0.001	0.009	0.17
HGDP Europe (N=159)	4	0	0.006	0.000	>1.5[b]

[a]Number of nonsynonymous and synonymous sites for PRRT2 are 668.5and 231.5, respectively.
[b]For the calculation of the pN/pS for Europe, the value of pS was set to 0.004 (the minimum of 1 synonymous variant observed in 159 individuals).

Table S1. Cohorts used in this study

	Sex	Asperger syndrome	Atypical autism	Autism	High functioning Autism	All ASD	Controls	HGDP
All *PRRT2* coding exons	Female	8	0	79	2	89	74	332
	Male	51	2	279	10	342	101	629
	All	59	2	358	12	431	186	961
A217PfsX8 screen only	Female	9	1	67	1	78	90	-
	Male	17	4	187	9	217	7	-
	All	26	5	254	10	295	92	-
A217PfsX8 screen All	Female	17	1	146	3	167	164	332
	Male	68	6	466	19	559	108	629
	All	85	7	612	22	726	278*	961

Table S2. Ethnicity of the cohorts

Ethnic	ASD			Controls
	All *PRRT2* coding exons	A217PfsX8 screen only	A217PfsX8 screen All	
Europe	261	213	474	278
Africa	40	6	46	-
Asia	13	2	15	-
Mixed origin	48	12	60	-
Unknown	69	62	131	-
All	431	295	726	278

Table S3. PRRT2 primers

Name	Chr	Position start	Position end	Amplicon size (bp)	Forward	Reverse	Temperature hybridation
PRRT2 exon 2 start	16	29731680	29732220	541	AACCCCAACTTTTCTTCCTCTC	CTCACTTCTGGTTTGGACACTG	55°C
PRRT2 exon 2 middle	16	29732078	29732695	618	GCTCCAGAAACCACAGAGACC	GATGGCAAGGATGATGTAGTCC	64°C
PRRT2 exon 2 end	16	29732390	29732912	523	TCTGAGAGTGTAGGGGAAAAGC	CTAGGGAGAGGCAAACAAAGG	59°C
PRRT2 exon 3 and 4	16	29733081	29733582	502	TTTCCACCTGATCCCTTCTG	CAGGCTCCCTTGGTCCTTAG	54°C

Table S4. *PRRT2* nonsynonymous variants identified in the HGDP

Mutation	North Africa N=29	Subsaharan Africa N=108	Middle East N=135	Europe N=159	Asia N=439	Oceania N=31	America N=64
S5N	0	0	0	0	1	0	0
P18T	0	0	0	0	1	0	0
E23K	0	0	0	1	0	0	0
P45S	0	2	0	0	0	0	0
P48R	0	0	0	0	1	0	0
S115K	0	0	1	0	0	0	0
P140A	0	0	0	1	43 heterozygotes 8 homozygote	3	10
D147H	0	0	0	0	3	0	0
T151N	0	0	0	0	1	0	0
P154S	0	0	0	0	5	0	0
E180K	0	0	1	0	0	0	0
A214P	0	0	0	0	5	0	1
P215R	0	0	0	0	4	0	0
P215T	0	0	0	0	1	0	0
P216L	1	0	0	1	1	0	0
R217Q	0	0	0	1	0	0	0
G241S	0	0	0	0	0	0	1
G258E	0	0	0	0	2	0	0
G258R	0	0	0	0	0	4 heterozygotes 1 homozygote	0
A272V	0	0	0	0	1	0	0
R311Q	0	0	0	0	1	0	0

Table S5. *PRRT2* synonymous variants identified in the HGDP

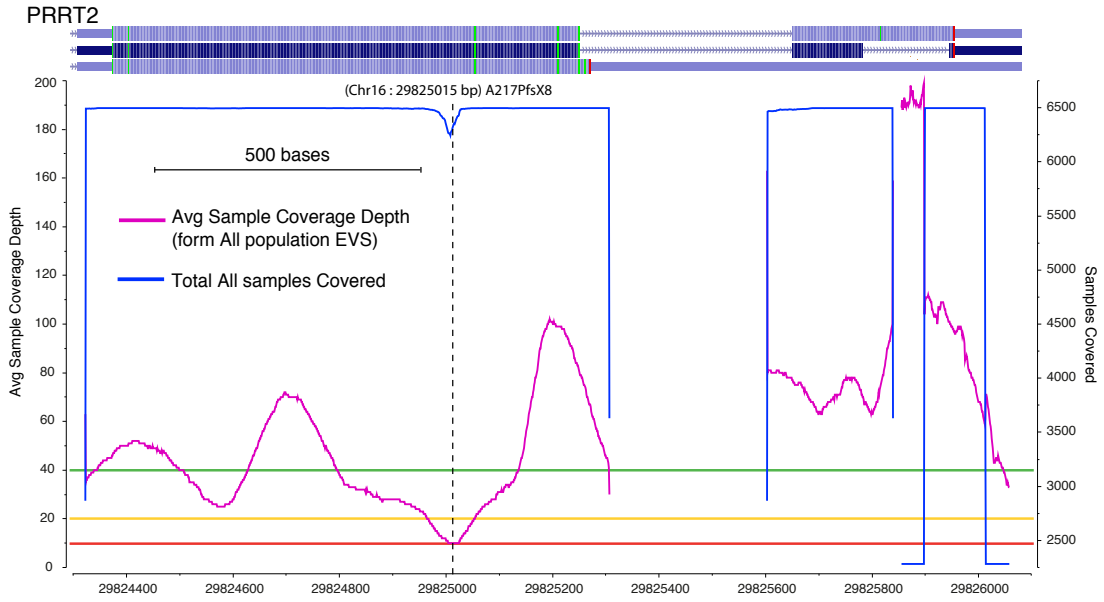
Cohort	Region	CEPH ID	Position	Variants	Allele
ASD Cohort study	Subsaharan Africa	PED-SAL-NGE-1013-005	16 : 29765126	L251L	C/T
		PED-SAL-LIO-1157-002	16 : 29765126	L251L	C/T
	Reunion	AU-RD-SAJ-209-004	16 : 29825914	P381P	A/C
	Europe	AU-GRE-SEB-132-004	16 : 29825026	R217R	A/G
HGDP	Subsaharan Africa	HGDP00460	16 : 29825020	P215P	C/T
		HGDP00465	16 : 29825020	P215P	C/T
		HGDP01411	16 : 29765126	L251L	C/T
		HGDP01419	16 : 29765126	L251L	C/T
		HGDP00920	16 : 29765126	L251L	C/T
		HGDP00927	16 : 29765126	L251L	T/T
		HGDP00932	16 : 29825203	C276C	C/T
	Asia	HGDP00146	16 : 29824756	E127E	G/A
	Middle East	HGDP00632	16 : 29765126	L251L	C/T
	North Africa	HGDP01271	16 : 29765126	L251L	C/T

Table S6. PRRT2 coding variants present in dbSNP, Exome variant Server (EVS) and the 1000 genome project (1KG)

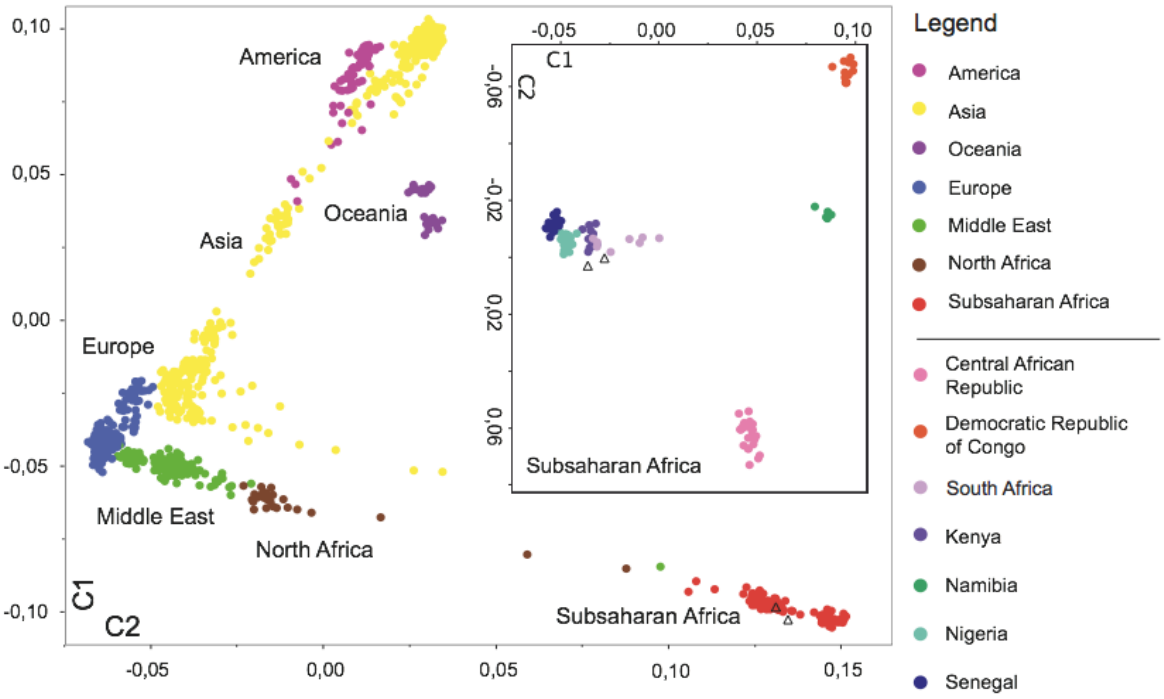
NCBI.37 Chr:position (bp)	rsID	Alleles	Avg Sample Read Depth	AA	cDNA Q7Z6L0-2 (bp)	Function GVS	Polyphen2	European American Genotype Count	African American Genotype Count	HGDP	Cohort	Source
16:29824442	rs140383655	A/G	49	E23K	67	missense	benign	AA=0/AG=16/GG=4284	AA=0/AG=3/GG=2194	HGDP	Cohorts	dbSNP_134
16:29824448	unknown	G/C	49	P25A	73	missense	benign	GG=0/GC=1/CC=4299	GG=0/GC=0/CC=2197	-	-	EVS
16:29824455	unknown	C/A	49	H27P	80	missense	possibly-damaging	CC=0/CA=1/AA=4299	CC=0/CA=0/AA=2197	-	-	EVS
16:29824508	rs11556732	T/C	40	P45S	133	missense	possibly-damaging	TT=0/TC=0/CC=4299	TT=0/TC=15/CC=2182	HGDP	-	dbSNP_120
16:29824599	rs147004110	T/C	27	P75L	224	missense	possibly-damaging	TT=0/TC=2/CC=4297	TT=0/TC=0/CC=2197	-	-	dbSNP_134
16:29824727	unknown	A/G	69	T118A	352	missense	benign	AA=0/AG=2/GG=4298	AA=0/AG=0/GG=2197	-	-	EVS
16:29824782	rs11556731	C/G		P136R	407	missense	benign			-	-	dbSNP_137
16:29824787	rs79182085	G/C	45	P138A	412	missense	benign	GG=0/GC=7/CC=4292	GG=0/GC=11/CC=2186	-	-	dbSNP_131
16:29824805	rs150501365	G/C	37	P144A	430	missense	benign	GG=0/GC=0/CC=4300	GG=0/GC=2/CC=2195	-	-	dbSNP_134
16:29824814	rs79568162	G/C		D147H	439	missense	benign			HGDP	Cohorts	-
16:29824875	unknown	T/C	29	T167I	500	missense	possibly-damaging	TT=0/TC=1/CC=4298	TT=0/TC=0/CC=2197	-	-	EVS
16:29824902	1KG_16_29824902	G/A		G176E	527	missense	benign			-	-	EVS, 1KG
16:29824955	rs139516010	A/G	23	E194K	580	missense	possibly-damaging	AA=0/AG=1/GG=4299	AA=0/AG=0/GG=2194	-	-	dbSNP_134
16:29825015	unknown	GCC/G/GC	10	-	640	Coding Complex	unknown	A1A1=8/A1A2=2/A1R=240/A2A2=21/A2R=501/RR=3204	A1A1=3/A1A2=1/A1R=149/A2A2=6/A2R=254/RR=1623	-	Cohorts	EVS
16:29825018	unknown	T/C	10	P215S	643	missense	probably-damaging	TT=0/TC=1/CC=4234	TT=0/TC=0/CC=2145	-	-	EVS
16:29825019	unknown	G/C	10	P215R	644	missense	probably-damaging	GG=0/GC=4/CC=4240	GG=0/GC=0/CC=2150	HGDP	-	EVS
16:29825022	rs76335820	T/C	10	P216L	647	missense	probably-damaging	TT=1/TC=60/CC=4195	TT=0/TC=11/CC=2156	HGDP	Cohorts	dbSNP_131
16:29825024	rs77838305	C/G		R217G	649	missense	probably-damaging	probably-damaging		-	-	-
16:29825025	rs75497546	G/A		R217Q	650	missense				HGDP	-	-
16:29825054	unknown	G/C	20	R227G	679	missense	probably-damaging	GG=0/GC=0/CC=4299	GG=0/GC=1/CC=2196	-	-	EVS
16:29825076	unknown	T/G	26	G234V	701	missense	probably-damaging	TT=0/TG=1/GG=4299	TT=0/TG=0/GG=2197	-	-	EVS
16:29825127	rs17850747	T/C		L251S	752	missense	possibly damaging			-	-	-
16:29825695	unknown	C/G	64	Q307H	921	missense	probably-damaging	CC=0/CG=1/GG=4299	CC=0/CG=0/GG=2197	-	-	EVS
16:29825949	rs144540943	C/T	99	V338A	1013	missense-near-splice	possibly-damaging	CC=0/CT=4/TT=4296	CC=0/CT=0/TT=2197	-	-	dbSNP_134

Table S7: Identity by state (IBS) values for the 2 individuals with ASD carrying the mutation p.A361_P362del compared with the IBS from different African populations.

Population	AU-RD-158-003			AU-RD-000-001		
	Min	Average	Max	Min	Average	Max
C.A.R	0.685	0.687	0.692	0.681	0.685	0.688
D.R. of Congo	0.678	0.680	0.683	0.675	0.676	0.677
Kenya	0.683	0.690	0.694	0.685	0.689	0.693
Namibia	0.672	0.674	0.676	0.668	0.670	0.671
Nigeria	0.690	0.694	0.697	0.690	0.692	0.694
Senegal	0.687	0.690	0.693	0.685	0.689	0.692
South Africa	0.689	0.693	0.695	0.689	0.692	0.696
AU-RD-158-003	-	-	-	0.694	0.694	0.694
AU-RD-000-001	0.694	0.694	0.694	-	-	-



Supplementary Figure1. Sequence coverage of *PRRT2* in Exome Variants Serveur. Total number of sample sequenced (in blue) and average read depth (purple) for PRRT2 in Exome Variant server. The nucleotide positions are according to PRRT2 from NCBI37/hg19. The average coverage of 10X, 20X and 40X are indicated in red, yellow and green respectively.



Supplementary Figure 2. Multiple dimension scale (MDS) of the genetic distance between individuals from the HGPD. Insert. MDS restricted to the African populations including the two individuals carrying the in frame deletion p.A361_P362del (triangles).

Discussion

La relation génotype-phénotype dans l'autisme et la dyslexie.

L'hétérogénéité génétique, la pénétrance incomplète et l'expressivité variable : Trois défis pour la génétique humaine

Les chercheurs et les cliniciens sont largement démunis face à l'hétérogénéité génétique des maladies qui limite très fortement l'identification des mutations causales. De plus, le paysage génétique et épigénétique de chaque individu influe sur la pénétrance et la sévérité des mutations. Même dans le cas de maladies apparemment monogéniques, les mutations peuvent avoir des conséquences variables selon les individus. Face à la difficulté à déterminer précisément la relation génotype-phénotype, on incrimine soit le fond génétique soit l'environnement sans réelle connaissance de l'effet de gènes « modificateurs » et des facteurs environnementaux en cause.

1.1 L'hétérogénéité génétique de l'autisme et la dyslexie

L'ensemble des données de la littérature ainsi que les résultats de ce travail démontrent clairement l'existence d'une très grande hétérogénéité génétique de l'autisme. Nous avons ainsi pu montrer que des gènes différents (et plusieurs voies biologiques différentes) pouvaient être en cause. Une analyse de la littérature et une estimation basée sur les études de CNVs et d'exomes suggèrent l'implication de plus de 500 gènes en cause dans l'autisme (Huguet, Ey et al. 2013). Cependant, il est actuellement très difficile de connaître le poids de chaque gène/mutation dans la vulnérabilité à l'autisme tant au niveau des populations atteintes (quels sont les gènes les plus mutés dans l'autisme ?) qu'au niveau individuel (quel est l'impact de cette mutation sur le risque de développer un autisme ?).

Pour la dyslexie, bien que peu (ou pas ?) de gènes aient été identifiés, il est probable qu'une grande hétérogénéité génétique soit retrouvée. En effet, les analyses de liaisons et de CNVs semblent confirmer cette hétérogénéité(Fisher and DeFries 2002, Scerri and Schulte-Korne 2010, Poelmans, Buitelaar et al. 2011). Il est important de noter que la dyslexie étant moins sévère que certaines formes d'autisme, on peut s'attendre à ce que la part des variants fréquents dans l'héritabilité soit plus importante pour la dyslexie que pour l'autisme. En effet, les variants de vulnérabilité à la dyslexie ont probablement un effet plus faible sur la condition physique de l'individu (« la fitness ») et peuvent donc passer de génération en génération sans être éliminés de la population.

Malgré cette très grande hétérogénéité génétique, il est certain que nous connaissons de mieux en mieux certaines voies biologiques qui semblent être responsables de l'autisme et dans une moindre mesure de la dyslexie. L'analyse des génomes complets va aussi permettre d'identifier de plus en plus de variants de vulnérabilité. Cependant, dans la très grande majorité des cas, la complexité d'un syndrome ne pourra pas être réduite à une simple mutation et nous devrons prendre en compte l'ensemble (ou du moins une partie pertinente) du génome de chaque individu ainsi que la part de l'environnement dans le développement de la maladie.

1.2 L'importance du fond génétique pour comprendre la pénétrance incomplète et l'expressivité variable

Lors de notre étude, nous avons pu noter l'importance des notions de pénétrance incomplète et d'expressivité variable. L'importance du fond génétique (Figure 1) a aussi été illustrée à plusieurs reprises.

Premièrement, nous avons pu confirmer que les mutations tronquantes de *PRRT2* n'étaient pas toujours pénétrantes. De plus, nous avons montré que la fréquence de mutations non synonymes *PRRT2* était très différente d'une population à l'autre. Ces résultats suggèrent qu'une mutation *PRRT2* dans un génome de composante ancestrale africaine a peut être plus de conséquences phénotypiques que dans un génome de composante asiatique ou européenne. Ceci expliquerait le très faible taux de variations non synonymes dans la population africaine comparée aux populations asiatiques ou européennes. Actuellement, nous ne pouvons pas conclure si cette différence de pression de sélection que nous avons observée est due au fond génétique de ces populations ou à leur environnement différent. Il est en effet concevable que la sévérité du phénotype associé aux mutations *PRRT2* soit modulée soit par le fond génétique (gènes suppresseurs) soit par l'environnement (diminution très importante de la « fitness » dues aux mutations *PRRT2* en Afrique par rapport à l'Asie ou l'Europe).

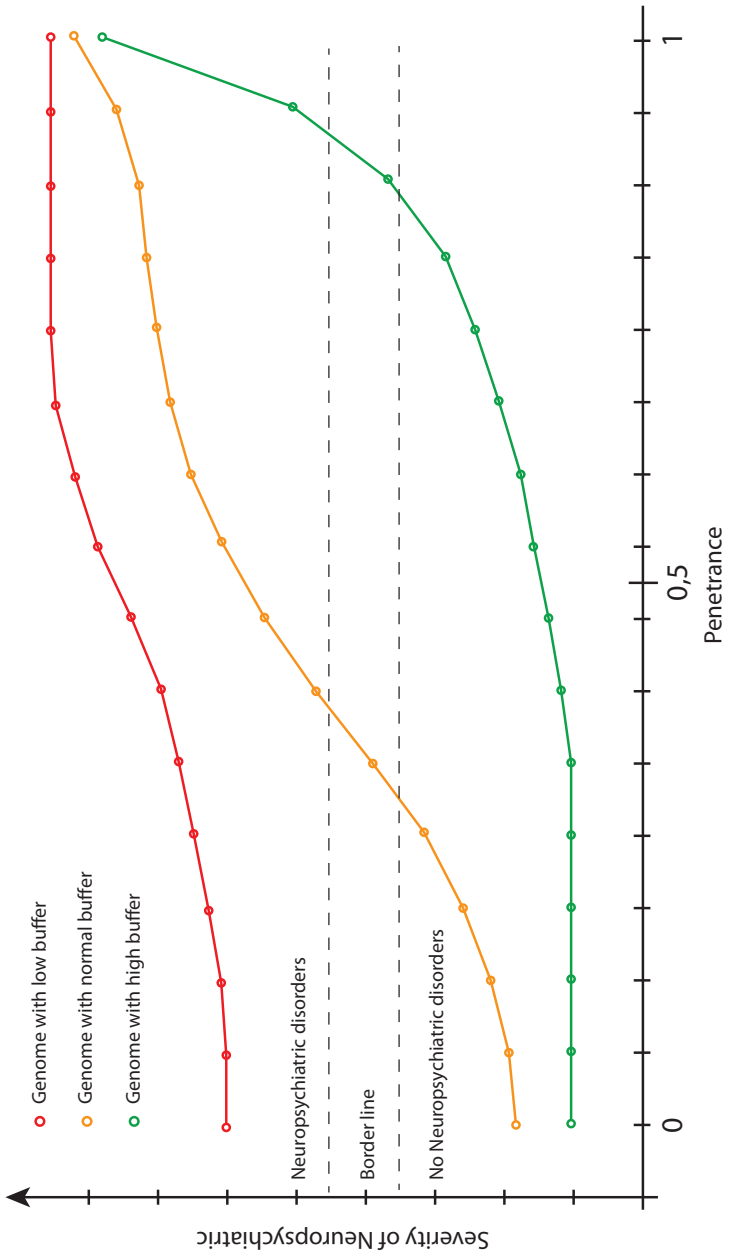


Figure 1. La pénétrance incomplète et l'impact des mutations rares sur la vulnérabilité aux troubles neuropsychiatriques. Selon le génome de l'individu, des mutations peuvent avoir un impact très différent. Le génome « rouge » a peu de possibilité de compenser (ou de « tamponner ») l'effet des mutations. La personne possédant ce génome aura un risque très élevé de développer une maladie psychiatrique. Au contraire, une personne possédant le génome « vert » sera protégé sauf pour certaines mutations avec une pénétrance très élevée. La personne avec le génome « orange » est intermédiaire.

Il est estimé que chaque individu possède 250 à 300 gènes ayant des mutations conduisant à une perte potentielle de fonction, ainsi que 50 à 100 mutations (à l'état hétérozygote) déjà associées à des maladies génétiques (Durbin, Abecasis et al. 2010). Des différences dans le type de sélection selon les populations ont déjà été décrites pour plusieurs gènes (Barreiro, Laval et al. 2008). L'exemple le plus connu est celui de l'allèle de l'hémoglobine S (HbS) qui cause la drépanocytose à l'état homozygote mais qui protège contre le paludisme à Plasmodium falciparum à l'état hétérozygote. La fréquence des mutations augmente donc davantage dans es populations en contact avec l'agent infectieux que dans les populations de pays où le paludisme n'est pas présent.

Deuxièmement, nous avons montré que des mutations génétiques probablement délétères étaient présentes chez des apparentés apparemment asymptomatiques : par exemple, les parents d'enfants autistes porteurs de délétion *NLGN1*, *IQSEC3*, et enfin l'individu 34 de la famille PL qui est porteur de l'haplotype à risque *CNTNAP2* et qui visiblement n'est pas dyslexique. Cependant, ces cas de pénétrance incomplète pourraient s'avérer être des cas avec expressivités variables. Pour répondre à cette question, il faudrait réaliser des examens complémentaires afin de déterminer si ces personnes apparemment asymptomatiques ont des phénotypes intermédiaires infra-cliniques, neurologiques, cognitifs ou psychiatriques. En effet, il a été montré que les apparentés de patients avec TSA avaient des scores plus élevés que la population générale pour les tests visant à détecter des traits autistiques(Constantino, Przybeck et al. 2000, Constantino and Todd 2003, Belmonte, Gomot et al. 2009).

Cette expressivité variable d'une mutation sur le niveau cognitif est très bien illustrée par l'étude de Moreno de Luca et al. (2013) sur les porteurs de délétions

22q11.2 ou 16p11.2(Moreno-De-Luca, Myers et al. 2013). Cette étude qui porte sur une grande cohorte de patients montre que les délétions ne donnent pas obligatoirement une déficience intellectuelle mais plutôt décalent le niveau de quotient intellectuel de 2 déviations standard par rapport à la population générale (Moreno-De-Luca, Myers et al. 2013). Ceci suggère donc que le QI de la personne va être influé à la fois par la présence de la mutation (-2DS) mais aussi très fortement par son fond génétique et son environnement. Il est intéressant de noter que dans notre étude sur l'autisme dans les îles Faeroe, l'individu porteur de la délétion *de novo* 22q11.2 est situé dans un cluster constitué uniquement de patients TSA (cluster B3). On pourrait donc faire l'hypothèse que la délétion 22q11.2 a déclenché l'autisme chez ce patient qui possèderait des allèles à risque regroupés au niveau de ce cluster. Dans un autre cluster, ce patient aurait pu présenter d'autres manifestations psychiatriques mais pas obligatoirement un autisme.

1.3 Les progrès et les défis dans l'analyse du génome humain

Afin de maîtriser l'hétérogénéité génétique et de comprendre les mécanismes en cause dans la pénétrance incomplète et l'expressivité variable, il est important de mieux définir ce que l'on appelle communément le fond génétique de l'individu. Actuellement, les puces à SNPs permettent de génotyper chez un seul individu des centaines de milliers voire plusieurs millions de SNPs. Grâce à ces technologies, nous avons la possibilité de génotyper la très grande majorité des variants fréquemment retrouvés dans les populations humaines en Europe. Par exemple, la dernière puce Illumina Omni 5 permet de génotyper ~ 4.5 millions de SNPs qui capturent 87% des SNPs connus avec une MAF >1%. De plus, cette information sur le fond génétique d'un individu va s'élargir dans le futur avec la possibilité d'avoir la totalité des variants localisés dans les exons voir la totalité des variants du génome d'un individu.

Cependant si l'on a maintenant la possibilité de détecter ces variants, les outils pour les annoter sont balbutiants. Les annotations fonctionnelles de ces variants sont très pauvres et pas toujours robustes. Plusieurs algorithmes, basés sur la conservation des variants au cours de l'évolution (score GERP) et sur la nature des changements protéiques (matrice de Grantham), peuvent nous aider à filtrer des variants délétères. Cependant, ces prédictions ne sont pas parfaites et les résultats peuvent être différents d'un algorithme à l'autre(Hao da, Feng et al. 2011). De plus, pour des traits comme l'autisme et la dyslexie qui sont définis par des capacités cognitives spécifiques à l'homme (ou du moins en apparence), il n'est pas certain que les variants en cause touchent des régions du génome conservées chez d'autres espèces.

Des scores d'haploinsuffisance sont aussi disponibles et basés sur l'occurrence de CNV dans la population générale et chez les patients, et peuvent être utilisés pour interpréter les résultats génétiques (Corpas, Bragin et al. 2012). Cette information est importante mais elle ne nous renseigne pas (ou très peu) sur la conséquence physiologique des mutations.

Enfin, la fonction des gènes est, dans la très grande majorité des cas, étudiée chez des organismes modèles comme *C. elegans, D. melanogaster*, ou *M. musculus* et le plus souvent dans un seul fond génétique et très homogène. Or on sait très bien que les mutations peuvent avoir des conséquences très différentes selon le fond génétique et selon les espèces. L'importance du fond génétique a été très joliment montrée dans une étude récente chez la drosophile. Les auteurs montrent que plusieurs types de mutations *scalloped* affectent la structure de l'aile très différemment selon la souche de Drosophila melanogaster (Chari and Dworkin 2013). Ils montrent aussi que les suppresseurs sont très différents d'une souche à l'autre avec des effets parfois opposés

(une mutation qui supprime l'effet de la mutation *scalloped* dans une souche va avoir au contraire un effet aggravant dans une autre souche).

Les conséquences de mutations sont aussi très différentes selon les espèces. Par exemple, il est très souvent observé que deux copies d'un gène doivent être mutées chez la souris pour avoir un phénotype alors que chez l'homme la perte d'une seule copie peut donner un phénotype très sévère. C'est le cas des mutations *SHANK2* ou *SHANK3* qui ne donnent pas de phénotype ou seulement très léger chez la souris hétérozygote (Peca, Feliciano et al. 2011, Schmeisser, Ey et al. 2012) alors que les délétions hétérozygotes causent un autisme chez l'homme (Durand, Betancur et al. 2007, Leblond, Heinrich et al. 2012).

1.4 Perspectives

Quelle que soit l'hétérogénéité génétique d'un trait et la nature des facteurs qui influent sur la pénétrance et l'expressivité variable, il nous semble maintenant crucial de mieux caractériser le fond génétique des patients et des témoins étudiés pour des traits complexes comme l'autisme et la dyslexie. Cette meilleure connaissance du « fond génétique » devrait permettre à la fois de mieux identifier les variants causaux impliqués dans les maladies et surtout de donner un pronostic plus précis sur les conséquences des mutations. Cette analyse du fond génétique peut être initiée dans des isolats génétiques comme dans notre étude sur les îles Faeroe et ensuite se généraliser à des populations plus hétérogènes sur le plan génétique. En particulier, l'utilisation des clusters basés sur la distance génétique entre individus (Identity By State) du génome entier (ou restreint à des loci d'intérêt) devrait permettre de mieux maitriser à la fois l'hétérogénéité génétique mais aussi la pénétrance incomplète et l'expressivité variable des mutations. Ainsi, il serait intéressant de faire une méta-analyse génétique des neuropathologies sur des traits quantitatifs, en regroupant les individus selon leur

distance génétique par rapport à leur génome ou pour un set de gènes (par exemple ceux exprimés dans le cerveau). Cela permettrait peut-être d'avoir la puissance statistique de détection pour les facteurs communs entre ces maladies et surtout de voir l'imbrication des gènes dans les maladies.

L'autisme, la dyslexie et *CNTNAP2*

Malgré la différence des atteintes cliniques, il est intéressant de noter que plusieurs arguments peuvent rapprocher les TSA, les troubles du langage et la dyslexie(Bishop 2003, Whitehouse, Watt et al. 2009, Gillberg 2010). Tout d'abord, les études sur les patients et les apparentés montreraient un continuum de dysfonctionnement linguistique entre les TSA, les troubles du langage et la dyslexie. Il a été observé dans plusieurs études que les parents ou frères et sœurs de personnes atteintes de TSA présentent des difficultés de langage (Kalnak, Peyrard-Janvid et al. 2012) bien que cela ne soit pas toujours retrouvé (Lindgren, Folstein et al. 2009).

Nos différentes études ont permis d'identifier de nouveaux gènes associés à l'autisme et de proposer de nouveaux gènes candidats confirmant le rôle des protéines synaptiques comme les protéines d'adhérence cellulaires (neuroligines, neurexines) et les protéines d'échafaudage (SHANK). Parmi tous ces gènes, *CNTNAP2* suscite un vif intérêt pour la communauté scientifique qui s'intéresse aux maladies neurologiques et psychiatriques comme les TSA et les troubles du langage(Newbury and Monaco 2010, State 2011, Penagarikano and Geschwind 2012, Graham and Fisher 2013, Rodenas-Cuadrado, Ho et al. 2013). *CNTNAP2* code pour une protéine neuronale d'adhérence (CASPR2) assimilée aux neurexines et qui est exprimée dans diverses régions cérébrales. Plusieurs arguments génétiques montrent que *CNTNAP2* est très

probablement un gène associé à plusieurs troubles neuropsychiatriques plus ou moins sévères (Figure 2).

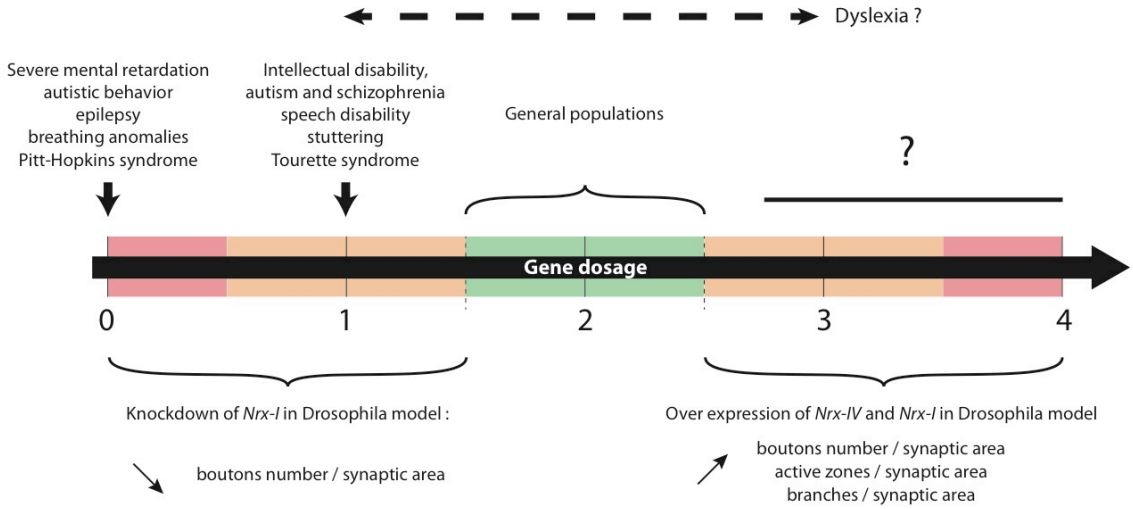


Figure 2. Les différentes maladies associées aux mutations CNTNAP2. Selon la dose génique de *CNTNAP2*, la sévérité de la maladie est différente. Il n'y a actuellement pas d'étude sur l'effet de la surexpression de *CNTNAP2*. Bien que dans le cas de mutations *FOXP1* ou *FOXP2*, qui sont des répresseurs de *CNTNAP2*, il est attendu d'avoir une surexpression du gène. Chez Drosophila Melanogaster, la réduction de *NrxIV* (l'orthologue de *CNTNAP2*) induit une réduction des boutons synaptiques. Sa surexpression entraine une augmentation des dendrites et de la densité synaptique(Zweier, de Jong et al. 2009).

2.1 Les variants rares de *CNTNAP2*

La première anomalie dans le gène *CNTNAP2* chez l'homme a été découverte en 2003 dans une famille avec un père souffrant de trouble obsessionnel compulsif (TOC) et chez deux enfants (un garçon et une fille) souffrant de plusieurs syndromes TOC, Gilles de la Tourette, déficience intellectuelle et retard de croissance. Les trois individus étaient porteurs d'une translocation dans le gène *CNTNAP2*(Verkerk, Mathews et al. 2003). La deuxième mutation a été retrouvée en 2006 dans une grande famille de la communauté Amish. Plusieurs membres de cette famille présentaient un syndrome d'épilepsie corticale avec dysplasie (CDFE syndrome) associé à une régression du langage, une légère macrocéphalie, une hyperactivité, un comportement impulsif et agressif et une diminution des réflexes ostéo-tendineux (Strauss, Puffenberger et al. 2006). L'analyse génétique a montré que 9 enfants étaient porteurs d'une mutation tronquante à l'état homozygote. Trois ans plus tard, Jackman et al. (2009) décrivent une fille de la même famille avec les même symptômes précédemment décrit mais porteuse d'une seule mutation à l'état hétérozygote(Jackman, Horn et al. 2009).

En 2008, 3 délétions hétérozygotes ont été identifiées chez 1 patient avec déficience intellectuelle et retard de langage et chez 2 patients indépendants atteints de schizophrénie et épilepsie(Friedman, Vrijenhoek et al. 2008). La même année, 3 études indépendantes simultanément publiées dans le même journal montrent une association entre *CNTNAP2* et l'autisme(Alarcon, Abrahams et al. 2008, Arking, Cutler et al. 2008, Bakkaloglu, O'Roak et al. 2008). La première étude de Alarcon et al. (2008) rapporte une analyse de traits quantitatifs (QTL) pour le langage sur le chromosome 7 chez 476 familles avec autisme (les 2 parents et un enfant). Un SNP rs2710102 centré sur *CNTNAP2* est significativement associé aux troubles du langage ($P<0.002$). La seconde étude de Arkin et al. (2008) rapporte à la fois une étude de liaison et une étude

d'association entre *CNTNAP2* et l'autisme. Les auteurs font une analyse de liaison avec 72 familles multiplexes (principalement des paires de germains) et montrent un lodscore > 3 centré sur *CNTNAP2.* Une analyse d'association chez 145 trios identifie ensuite rs7794745 comme étant associé à l'autisme (P<0.001). De façon intéressante, les associations étaient surtout positives pour les familles possédant uniquement des garçons atteints. Cependant cette différence est peut être due à un nombre moins important de filles et donc à une puissance statistique plus faible pour détecter une association. Dans la dernière étude de Bakkaloglu et al. (2008), 635 patients et 942 témoins ont été séquencés pour les exons de *CNTNAP2*. Les auteurs ont identifié 27 variants non synonymes dont 13 étaient uniques chez les patients (8 prédits comme délétères). Un variant I869T a été transmis à 4 enfants dans 3 familles différentes mais n'a pas été retrouvé chez 4010 témoins.

En 2009, Zweier et al. (2009) ont identifié 3 patients dans 2 familles indépendantes avec des mutations sur les 2 allèles de *CNTNAP2* (Zweier, de Jong et al. 2009). Deux frères et sœurs sont hétérozygotes composites pour 2 délétions exoniques différentes. L'autre patient est porteur d'une délétion sur un allèle et d'une mutation d'épissage sur l'autre allèle. Ces patients sont autistes avec déficience intellectuelle, épilepsie et troubles respiratoires. Ces symptômes ressemblent à ceux observés dans le syndrome de Pitt-Hopkins. Dans la même étude, les auteurs présentent 2 autres patients avec des symptômes similaires et porteurs de mutation *NRXN1* sur les deux allèles. Comme Jackman et al. (2009) décrivant une patiente hétérozygote de la famille Amish qui présentait des symptômes uniquement observés chez les patients homozygotes (Jackman, Horn et al. 2009), la même équipe identifie en 2011 des patients hétérozygotes mais sévèrement atteints et présentant les symptômes majeurs du syndrome de Pitt-Hopkins (Gregor, Albrecht et al. 2011).

En 2010, Elia et al. (2010) étudient les CNVs de 335 patients avec hyperactivité et identifient une délétion intronique chez 1 patient (Elia, Gai et al. 2010). Mefford et al. (2010) étudient les CNVs chez 517 patients avec épilepsie et trouve 1 patient porteur d'une délétion (Mefford, Muhle et al. 2010). Depuis, de nombreux cas de remaniements de *CNTNAP2* ont été décrits avec translocations et remaniements du génome chez des patients atteint de TSA avec troubles du langage (Rossi, Verri et al. 2008, Poot, Beyer et al. 2010, Mikhail, Lose et al. 2011, Al-Murrani, Ashton et al. 2012), mais aussi dans un cas de bégaiement (Petrin, Giacheti et al. 2010). Un cas de dyslexie avec une duplication intragénique de CNTNAP2 a aussi été identifié (Veerappa, Saldanha et al. 2013). Néanmoins, dés 2007, plusieurs cas de pénétrance incomplète ont été rapportés pour des porteurs de translocations ou de mutations hétérozygotes (Belloso, Bache et al. 2007).

Les analyses d'exomes, dès 2011, ont aussi permis d'identifier certains variants rares de *CNTNAP2*(O'Roak, Deriziotis et al. 2011, Worthey, Raca et al. 2013). L'étude de Roak et al, 2011, observe dans une famille la présence de deux variants : un *de novo* dans *FOXP1* (A339SfsX4) et le second hérité de la mère dans *CNTNAP2* (H275A). Par la suite, les auteurs ont testé l'expression de *CNTNAP2* en présence de la forme sauvage ou en présence de la forme tronquée de *FOXP1* dans des cellules HEK293T. Ils ont ainsi montré que par rapport à la protéine sauvage FOXP1, la forme tronquée augmente l'expression de *CNTNAP2*. Ceci est en accord avec l'existence d'un effet régulateur négatif de FOXP1 sur *CNTNAP2* comme cela avait déjà été montré pour FOXP2(Vernes, Newbury et al. 2008). L'étude de Wortheu et al (2013), porte sur une cohorte de 10 patients âgés de 3 à 19 ans avec apraxie de la parole (« Childhood apraxia of speech »). Il s'agit d'un syndrome pédiatrique rare, grave et persistant au niveau du langage moteur mais aussi avec des déficits sensori-moteurs et cognitifs. Ainsi sur les 10 patients, 3 ont une

mutation dans *CNTNAP2* et un patient a une mutation dans *FOXP1*. Tous ces patients présentent une déficience intellectuelle et un TSA en plus de l'apraxie de la parole. Ces observations renforcent l'implication de *CNTNAP2* dans différentes neuropathologies y compris les troubles du langage. Enfin très récemment, l'étude par liaison génétique d'une famille consanguine égyptienne a permis d'identifier chez des patients souffrant d'épilepsie une mutation homozygote dans CNTN2, la protéine partenaire de CNTNAP2 (Stogmann, Reinthaler et al. 2013).

La majorité de ces résultats est basée sur l'observation de variants rares ou *de novo*, mais on peut aussi se poser la question des variants fréquents de *CNTNAP2* et de leurs conséquences phénotypiques dans la population générale.

2.2 Les études d'association avec *CNTNAP2*

Comme pour tout gène très médiatisé, une multitude d'analyses ont été entreprises pour démontrer l'implication de *CNTNAP2* dans différentes maladies (Li, Hu et al. 2010)-(Wang, Liu et al. 2010, Newbury, Paracchini et al. 2011, Sharma, Gao et al. 2011, Stein, Yang et al. 2011, Whitehouse, Bishop et al. 2011)-(Peter, Raskind et al. 2011)-(Ji, Li et al. 2013)-(Toma, Hervas et al. 2013) ou trait de personnalité(Terracciano, Sanna et al. 2010).

Concernant les troubles autistiques, les premiers résultats positifs d'association proviennent de deux études publiées en 2008(Alarcon, Abrahams et al. 2008, Arking, Cutler et al. 2008). Il est important cependant de rappeler que ces études ont été effectuées sur un nombre limité de patients (476 pour Alarcon et al. (2008) et 145 trios pour Arkin et al. (2008)) et qu'elles ne détectaient pas le même SNP associé (rs2710102 pour Alarcon et al. (2008) et rs7794745 pour Arkin et al. (2008)). Depuis, deux grandes études sur un nombre beaucoup plus élevé de patients ont été réalisées. La première de

Anney et al. (2012) comporte une étude pangénomique d'association (GWAS) sur 2705 familles avec TSA(Anney, Klei et al. 2012). Les auteurs ne trouvent pas d'association significative de SNPs sur l'ensemble du génome, mais cependant un des SNPs dont l'association aux troubles autistiques est la plus forte se trouve localisé dans CNTNAP2 (rs1718101). Une restriction de la cohorte aux patients européens sans déficience intellectuelle a permis d'augmenter la significativité (P-value = $7{,}8 \times 10^{-9}$; OR = 2,13). Ces résultats étaient très encourageants, mais ceux d'une seconde étude portant dans une première phase sur 408 trios et 323 familles multiplexes et dans une deuxième phase sur 2051 familles (572 multiplexes et 1479 trios) sont plus décevants (Sampath, Bhat et al. 2013). En effet, si des variants sont associés aux TSA dans la première phase (rs17170073, $p=2 \times 10^{-4}$; rs2215798, $p=1{,}6 \times 10^{-4}$; rs2710093, $p=9 \times 10^{-6}$; rs2253031, $p=2{,}5 \times 10^{-5}$), aucun n'est trouvé significativement associé dans la seconde.

Ces résultats montrent très vraisemblablement que si des SNPs de *CNTNAP2* sont associés à l'autisme, ils ne confèrent qu'un très faible risque individuel. Ainsi, un grand nombre de patients et d'individus témoins seraient absolument nécessaires à la détection d'une association statistiquement significative. Mais surtout, ces résultats pointent un très grand problème méthodologique. En effet, ils posent la question de la pertinence de telles études d'association qui considèrent l'autisme comme un trait binaire (atteint *vs* non atteint) et qui négligent toute l'hétérogénéité de ce trouble complexe.

Plusieurs études ont tenté de mettre en évidence une association entre *CNTNAP2* et des traits intermédiaires entre les gènes et les données cliniques comme des mesures d'imagerie cérébrale ou de potentiels évoqués cérébraux. Un ou plusieurs SNPs est donc génotypé et utilisé pour stratifier les individus qui sont étudiés. Ainsi, il a été rapporté que chez des individus de la population générale, les porteurs homozygotes de « l'allèle

à risque » de rs7794745 présentaient des différences anatomiques dans différentes régions du cerveau comme une réduction dans la matière grise et blanche au sein du cervelet, du gyrus fusiforme, et du cortex frontal. (Tan, Doke et al. 2010). Des analyses d'imagerie fonctionnelle ont aussi montré des défauts de connectivité à longue distance (Scott-Van Zeeland, Abrahams et al. 2010, Dennis, Jahanshad et al. 2011, Clemm von Hohenberg, Wigand et al. 2013) et des anomalies lors du traitement linguistique (augmentation de l'activation dans le gyrus frontal inférieur droit et dans le cortex temporal droit) (Whalley, O'Connell et al. 2011) pour les porteurs de l'allèle à risque de rs2710102. Enfin, une analyse de mesure de potentiels évoqués montre également des différences lors de l'écoute et du traitement linguistique de phrases mais cette fois ci avec le SNP rs7794745 (Kos, van den Brink et al. 2012).

Si ces analyses apparaissent attractives du fait qu'elles peuvent potentiellement donner des indications plus précises sur le rôle de CNTNAP2, il est important de noter qu'elles sont effectuées sur un nombre relativement faible d'individus (N=32-314) et qu'il serait absolument nécessaire de les répliquer. Il est important de noter qu'aucune étude montrant l'absence d'association sur l'analyse des variants *CNTNAP2* n'a été publiée à ce jour. Cette absence de résultats négatifs pour les études d'associations est bien connue et rend difficile les méta-analyses.

L'une des perspectives de notre étude sera d'étudier par l'imagerie cérébrale le cerveau des porteurs et de non porteurs de l'haplotype à risque de la famille PL. Même si actuellement nous ne connaissons pas le ou les variants causaux, il serait très intéressant de connaître les différences anatomiques et fonctionnelles associées à cet haplotype à risque et de voir s'il est similaire à ce qui a été décrit dans les analyses sur les variants fréquents de *CNTNAP2*.

2.3 L'expression et la régulation de CNTNAP2

CNTNAP2 est le plus long gène du génome humain (2.3Mb) et contient 24 exons (Figure 3)(Nakabayashi and Scherer 2001). La protéine CNTNAP2 possède plusieurs domaines protéiques : 2 domaines EGF (epidermal grow factor), 4 domaines laminine G, 1 domaine FA5/8 de type C (discoidin), 1 domaine « fibrinogen-like », 1 domaine « thrombospondin N-terminal-like», 1 domaine « 4.1 binding domain », 1 domaine transmembranaire et un 1 site de liaison PDZ.

Plusieurs mRNAs sont référencés dans GenBank (issus de promoteurs alternatifs), mais seulement deux isoformes sont décrites dans Refseq. Ces deux isoformes sont très différentes car la plus grande contient tous les domaines protéiques alors que la plus petite ne contient que le domaine transmembranaire, les domaines de fixation aux protéines 4.1 et aux domaines PDZ.

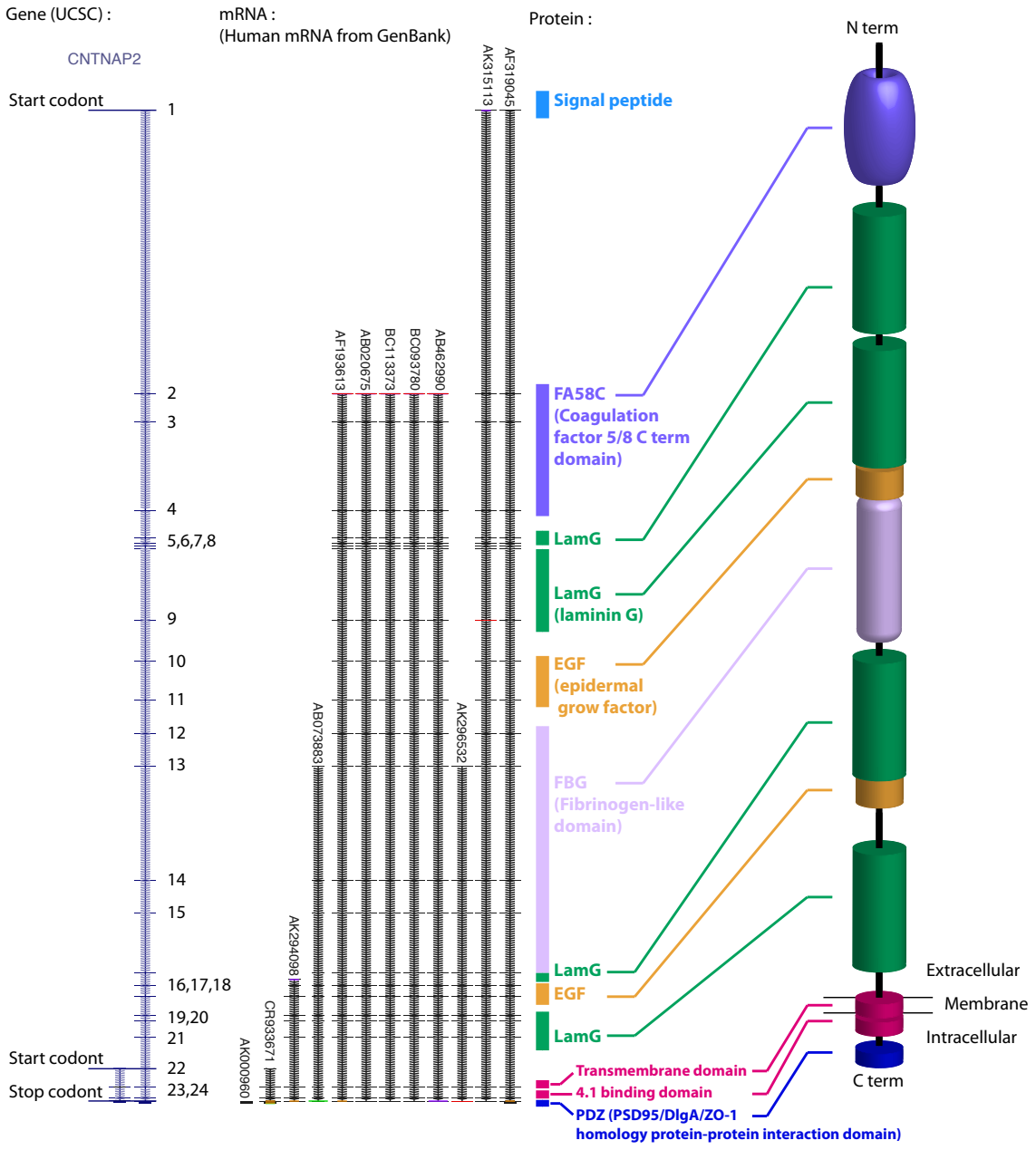


Figure 3. La structure génomique et protéique de CNTNAP2. Schéma de la structure du gène CNTNPA2 et de ces différentes isoformes répertorié dans GeneBank, ainsi que la description des domaines protéiques.

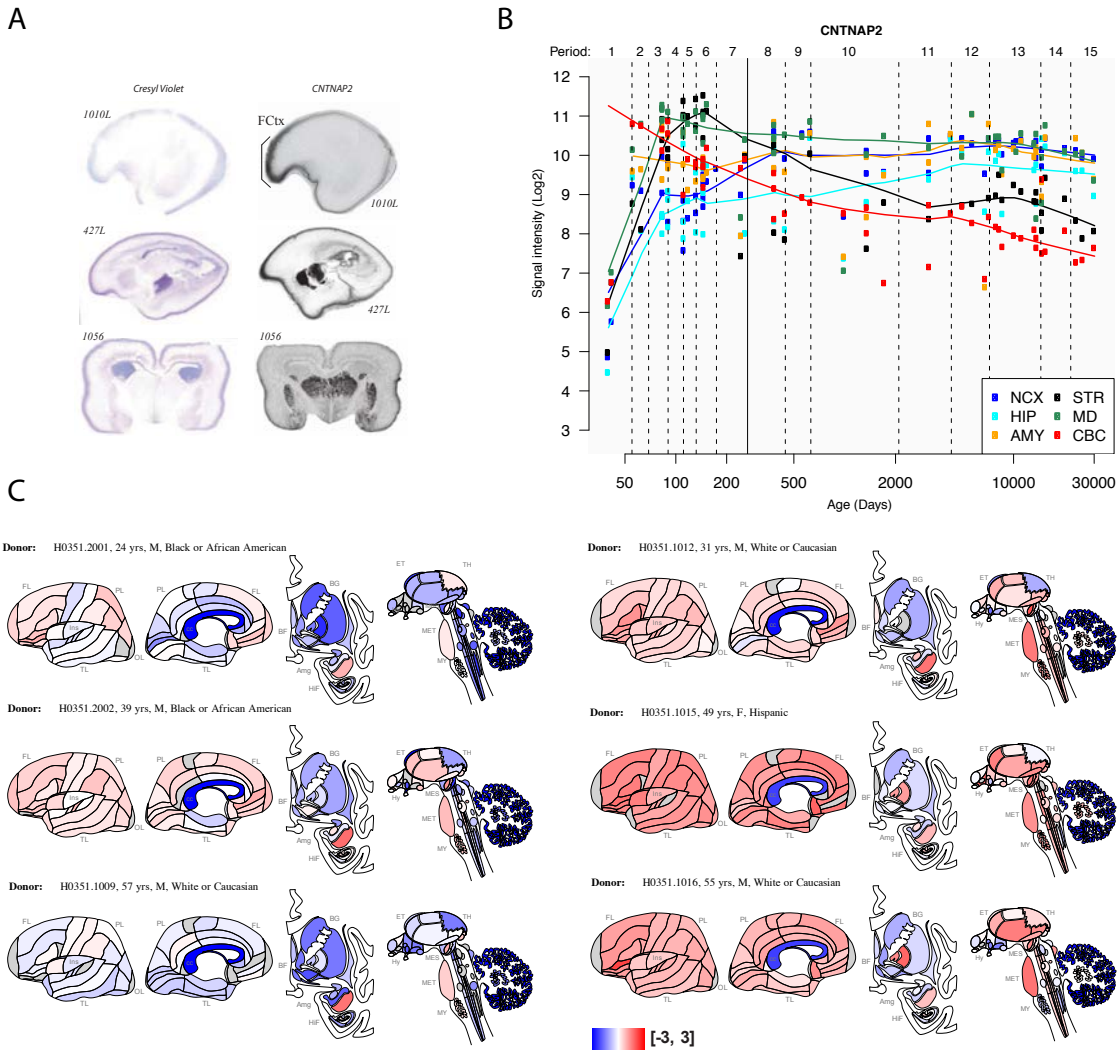


Figure 4. Expression de *CNTNAP2* chez l'homme. (a) Hybridation in situ de *CNTNAP2* dans le cerveau d'humain de 19-20 semaines de gestation sur coupe sagittal (1010L et 427L) et coronale (1056 et 1110). Les données sont issus de Abrahams et al (2007). (b) Graphique issu du projet « human brain transcriptome » (HBT, http://hbatlas.org/) montrant l'expression au cours du temps réalisé sur 1340 échantillons de tissus post-mortem prélevés à partir de 57 individus. (c) Illustration schématique des niveaux de transcrits *CNTNAP2* sur 6 cerveau adultes. Les données sont issues de la base de données Allen Brain (http://www.brain-map.org/).

L'expression de *CNTNAP2* est essentiellement localisée dans le cerveau mais aussi dans les tissus nerveux périphériques comme la moelle épinière(Poliak, Gollan et al. 1999). En 2007, Abrahams *et al.* publient leur travaux sur la comparaison d'expression entre deux aires cérébrales chez l'homme qui sont le gyrus temporal supérieur et le cortex sur des cerveaux de 18,6 semaines de gestation (Abrahams, Tentler et al. 2007). Ils ont ainsi défini 343 gènes avec des expressions différentes entre les deux aires cérébrales, dont 44 qui sont plus exprimés dans le gyrus temporal supérieur et 299 plus exprimés dans le cortex avec *CNTNAP2*. Par la suite ils ont observé par hybridation *in situ* certains transcrits dont *CNTNAP2* qui était très fortement exprimé dans le cortex frontal ainsi que dans plusieurs structures sous-corticales, aux niveaux du thalamus dorsal, du noyau caudé, du putamen et de l'amygdale (Figure 4).

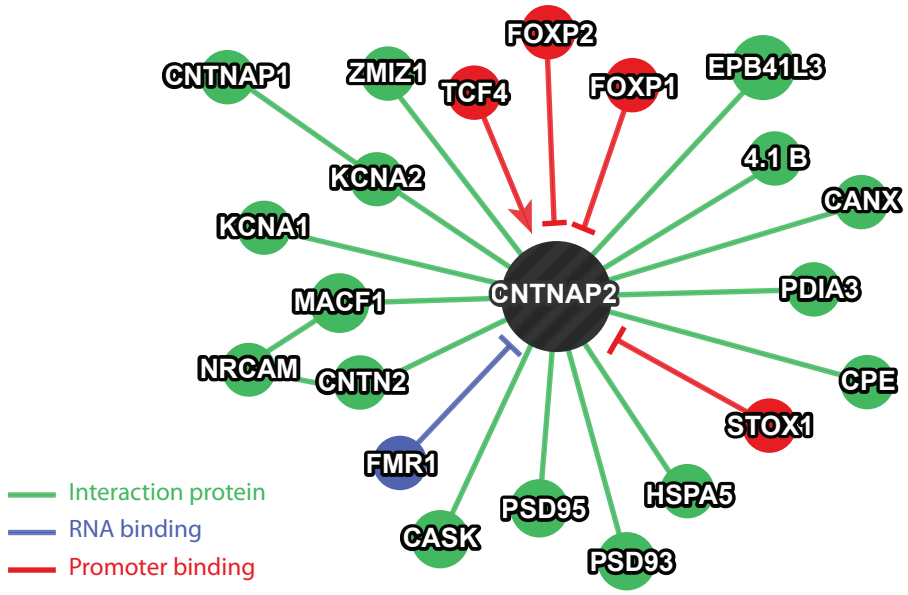


Figure 5. Interactome de CNTNAP2. Représentation des régulations de *CNTNAP2* par les facteurs de transcription TCF4, FOXP1, FOXP2 et STOX1, ainsi qu'au niveau traductionnel avec FMRP. Puis les interactions protéine-protéine entre CNTNAP2 et KCNA1 et KCNA2, ZMIZ1, EPB41L3, CANX, PDIA3, CPE, HSPA5, CASK, PSD93, PSD95, Protéine 4.1, CNTN2 et MACF1.

Lorsque par là suite ils ont comparé aux rongeurs (rat et souris) au même stade de développement, ils remarquent que l'expression dans le cortex frontal est spécifique de l'homme. A la vue de ces résultats, les auteurs posent l'hypothèse que cette expression différentielle de *CNTNAP2* pourrait être un des éléments qui joue un rôle dans les capacités cognitives spécifiques à l'homme.

Plus récemment en 2011, le données du projet « human brain transcriptome » (HBT) montrent que l'expression de *CNTNAP2* est présente précocement dans le développement embryonnaire et semble se maintenir au cours du développement de l'individu même si l'expression de celle-ci semble diminuée dans le cervelet et striatum ce qui n'est pas le cas de l'hippocampe, du néocortex, de l'amygdale et noyau du médiodorsal du thalamus (Kang, Kawasawa et al. 2011).

Même si il est trop tôt pour faire une analogie entre les réseaux neuronaux impliqués dans le chant chez l'oiseau et dans la parole chez l'homme, il est intéressant de noter que chez le Diamant Mandarin l'expression de *CNTNAP2*, comme pour *FOXP2*, est plus intense au niveau des noyaux de l'arcopallium qui sont associés au chant et homologue à la couche 5 du cortex moteur chez l'homme(Panaitof, Abrahams et al. 2010, Condro and White 2013). En particulier, son expression était plus intense chez les mâles en accord avec le dimorphisme sexuel du chant dans cette espèce(Panaitof, Abrahams et al. 2010, Condro and White 2013).

Plusieurs études ont tenté de caractériser les facteurs de régulation de *CNTNAP2*. Ici, il est intéressant de rappeler que CNTNAP2 est le plus grand gène du génome et que l'on peut estimer que la transcription d'un ARNm prendrait environ 20h en considérant une vitesse de transcription ± 20 nucleotides par secondes (Ucker and Yamamoto 1984). La transcription de *CNTNAP2* semble être régulée par FOXP2, FOXP1, TCF4 et STOX1

(Figure 4). L'étude de Vernes *et al.* en 2008 montre que FOXP2 régule négativement *CNTNAP2* (Vernes, Newbury et al. 2008). Les auteurs décrivent aussi par « chromatin immunoprecipitation » l'un des site de fixation de FOXP2 au niveau de l'intron 1 de *CNTNPA2*. De plus dans le cortex cérébral, Ils observent une co-expression de FOXP2 et CNTNAP2 au niveau des couches II et III du cortex. Enfin, Roak et al, 2011, montre que, comme FOXP2, FOXP1 régule aussi négativement expression de *CNTNAP2* (O'Roak, Deriziotis et al. 2011).

En 2012, Van Abel D et al (2012) rapportent par « chromatin immunoprecipitation » des cibles de STOXIA dont *CNTNAP2* (van Abel, Michel et al. 2012). Ils montrent aussi que *CNTNAP2* a une expression diminuée dans l'hippocampe des patients atteints de la maladie d'Alzheimer où l'expression *STOX1A* est augmentée. Ces résultats suggèrent que, comme FOXP1 et FOXP2, STOX1A régule négativement *CNTNAP2*. Par contre, la même année, Forest *et al.* montrent grâce à une méthode de « luciferase reporter assay » que TCF4, qui par ailleurs est muté dans le syndrome de Pitt-Hopkins (Rosenfeld, Leppig et al. 2009), régule positivement l'expression de *CNTNAP2* (Forrest, Chapman et al. 2012).

Au niveau de sa traduction, il semblerait que les ARNm CNTNAP2, comme beaucoup d'autres transcrits de gènes associés à l'autisme (Darnell, Van Driesche et al. 2011), soit une des cibles de FMRP (Ascano, Mukherjee et al. 2012). Il est donc possible que via la régulation par FMRP, les ARNm de CNTNAP2 soient traduits en fonction de l'activité neuronale, mais ceci reste à être démontré.

2.4 Les multiples fonctions de *CNTNAP2*

CNTNAP2 (appelée aussi NrxIV ou Caspr2) a été découverte en 1996 chez Drosophila melanogaster pour son rôle majeur dans l'établissement de la formation et la fonction de

la barrière hémato-encéphalique(Baumgartner, Littleton et al. 1996). Les embryons mutants sont paralysés et les études électrophysiologiques indiquent qu'il y a une rupture de la barrière hémato-encéphalique. Les auteurs montrent ainsi le premier rôle d'interaction cellule-cellule de *CNTNAP2*. En 1999, Poliak *et al.* identifient l'orthologue chez le rat et montrent sa localisation au niveau des nœuds de Ranvier des axones myélinisés (Poliak, Gollan et al. 1999). Alors que la protéine CNTNAP1 (ou Caspr ou Paranodin) est présente aux jonctions paradonales (Einheber, Zanazzi et al. 1997, Menegoz, Gaspar et al. 1997, Peles, Nativ et al. 1997), CNTNAP2 est très précisément co-localisée avec les canaux potassiques aux niveaux des régions juxtaparanodales (Figure 6). D'ailleurs CNTNAP2 s'associe directement au niveau de sa partie C terminale et de son domaine de fixation PDZ avec les canaux Kv1.1, KV1.2 et KVbeta2. Par la suite de nombreux papiers ont confirmé la localisation de CNTNAP2 au niveau des nœuds de Ranvier et de sont interaction avec les canaux potassique, la protéine 4.1 et les protéines d'échafaudage PSD95 et PSD93(Arroyo, Xu et al. 2001, Rasband, Park et al. 2002, Horresh, Poliak et al. 2008, Horresh, Bar et al. 2010, Duflocq, Chareyre et al. 2011). En 2003, deux équipes montre le rôle essentiel de l'association entre la protéine contactine CNTN2 (appelée aussi TAG1 ou AXONIN 1) et CNTNAP2 pour l'organisation en cluster des récepteurs potassique au niveau des régions juxtaparanodal des fibres myélinisés (Poliak, Salomon et al. 2003, Traka, Goutebroze et al. 2003). Poliak et al. (2003) montrent une désorganisation des canaux potassiques aux niveau de ces régions chez la souris invalidée pour CNTNAP2 (Poliak, Salomon et al. 2003). Traka et al. (2003) montrent que chez la souris invalidée pour CNTN2, CNTNAP2 ne s'accumulent plus au niveau des régions juxtaparanodes, et l'enrichissement normal des canaux K + dans ces régions est fortement perturbé, à la fois dans le système nerveux central et périphérique. De plus ils montrent que CNTN2 est exprimée dans les neurones et

cellules gliales, et associe en cis avec CNTNAP2 au niveau de l'axone et en trans avec elle même entre la cellule gliale et le neurone.

Enfin, une étude récente de Anderson et al. (2012) nous renseigne sur une autre fonction de CNTNAP2 dans l'assemblage des circuits neuronaux. Leur étude porte sur l'observation de culture de neurones corticaux de souris transfectés par un vecteur lentiviral contenant un shRNA qui diminue l'expression de *CNTNAP2*. Cette diminution de *CNTNAP2* entraîne une baisse de l'activité des réseaux neuronaux, de l'arborisation dendritique et du développement des épines dendritique. Ceci est accompagné d'une modification morphologique avec une diminution globale du nombre de synapses excitatrices et inhibitrices. Ces résultats obtenus chez la souris confirment le rôle synaptique de CNTNAP2 déjà rapporté chez *Drosophila melanogaster*, ou la surexpression de *Nrx-IV* qui est l'orthologue de *CNTNAP2* induit une augmentation du nombre de boutons synaptiques et des dendrites (Zweier, de Jong et al. 2009).

Au vu de l'ensemble de ces résultats, CNTNAP2 remplirait donc une fonction au niveau de l'organisation des circuits neuronaux en plus de sa fonction dans l'organisation de nœud de Ranvier.

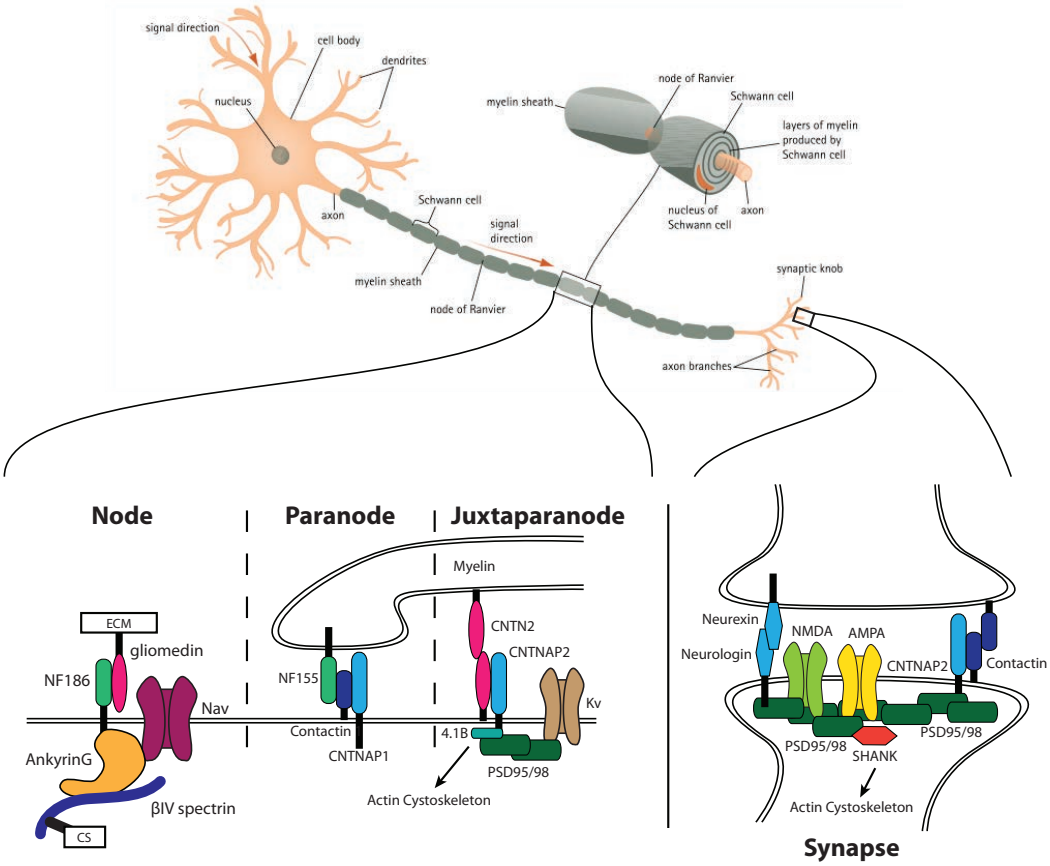


Figure 6 : Schéma de la localisation de CNTNAP2 au niveau des nœuds de Ravier et à la synapse.

2.5 Les conséquences possibles du dysfonctionnement de CNTNAP2

Les rôles de CNTNAP2 sont multiples puisque cette protéine est présente à la fois au niveau des nœuds de Ranvier mais aussi au niveau synaptique. En 2011, Olga Penagarikano et al . étudient la souris invalidée pour *CNTNAP2* générée par Poliak et al. (2003)(Poliak, Salomon et al. 2003) et observent une certaine similitude avec les symptômes cliniques des patients porteurs de mutations CNTNAP2. Entre autres, les souris présentent une épilepsie, des troubles des interactions sociales, des stéréotypies et de l'hyperactivité (Penagarikano, Abrahams et al. 2011). L'analyse du cerveau montre des ectopies neuronales au niveau du corpus callosum, suggérant des anomalies de

migration de neurones. Dans le cortex somatosensoriel, il y a un enrichissement de neurones dans les couche V et VI, un appauvrissement dans la couche II et une diminution des interneurones GABAergic. Ces anomalies sont accompagnées d'une diminution du niveau de synchronisation des neurones dans le cortex somatosensoriel. Ainsi, l'absence de CNTNAP2 a plusieurs conséquences à la fois sur la migration cellulaire, sur le nombre d'interneurones GABAergic et sur la synchronisation des neurones. On peut donc supposer que des anomalies similaires pourraient être présentes chez les patients. Or il semble qu'effectivement ce soit le cas.

Chez les patients d'origine Amish et porteurs d'une mutation homozygote, les données d'imagerie montrent une augmentation de volume au niveau du lobe temporal droit ainsi qu'un amincissement de la matière grise corticale et un flou à la jonction entre la matière grise et la matière blanche (Strauss, Puffenberger et al. 2006). Trois patients ont subi une chirurgie afin de contrôler leurs crises d'épilepsie qui malheureusement réapparaitront 6 à 15 mois après la chirurgie. L'analyse des prélèvements a permis de montrer pour les 3 enfants, des anomalies similaires dans l'organisation et la densité des neurones au niveau du cortex et de l'hippocampe. Les neurones étaient en plus grand nombre et tassés en paquet dans des colonnes. Certains neurones avaient une forme ronde au lieu de la forme pyramidale normale. Certains avaient même deux noyaux et une organisation dendritique anormale, avec une orientation inappropriée. De nombreux neurones étaient localisés au mauvais endroit (ectopie) et avaient peuplé la matière blanche sous corticale. Tous ces éléments suggèrent fortement des anomalies de la migration neuronale. Chez aucun des enfants, il n'y avait d'évidence de néoplasme, d'inflammation ou de dysplasie vasculaire.

Si l'on reprend l'ensemble de ces données, il semblerait que l'absence de CNTNAP2 cause une altération des réseaux neuronaux glutamatergiques et

GABAergiques et une désynchronisation neuronale. Ces anomalies pourraient alors influer sur la migration des neurones au cours du développement et induire les ectopies observées à la fois chez la souris et l'homme. Ce problème d'ordre développemental pose ainsi la question de la réversibilité de telles anomalies chez les enfants et les adultes porteurs des mutations CNTNAP2 (Delorme, Ey et al. 2013). Un résultat encourageant obtenu chez la souris invalidée pour Cntnap2 est l'effet positif du traitement par le risperidone qui semble améliorer une partie des symptômes chez la souris comme les comportements stéréotypés(Penagarikano, Abrahams et al. 2011).

2.6 Vers la confirmation d'un défaut de migration neuronal pour un sous groupe de patients atteints d'autisme ou et dyslexie ?

Plusieurs fois au cours de cette thèse, nous avons croisé le chemin de *CNTNAP2*. Ce gène est connu pour l'autisme et nous montrons qu'il est aussi en cause très probablement dans la dyslexie. De plus, même si nos résultats ne sont que préliminaires, nous avons aussi détecté une association significative entre des SNPs de *CNTNAP2* et l'autisme chez un sous-groupe de patients des îles Faeroe.

La piste déjà proposée des anomalies de migration neuronale dans la dyslexie et l'autisme pourrait donc se confirmer avec nos résultats qui impliquent *CNTNAP2* dans ces deux syndromes. Néanmoins au vu des résultats sur les altérations produites par les mutations *CNTNAP2* au niveau du cerveau chez l'homme et les modèles animaux, d'autres pistes subsistent, comme des anomalies de la myélinisation des axons, de la synchronisation des neurones. Il est trop tôt pour établir sur la base de résultats purement génétiques des hypothèses solides sur les mécanismes physiologiques et le rôle de CNTNAP2. Néanmoins, nous pouvons proposer plusieurs pistes pour des études ultérieures.

Il faudrait que nous puissions investiguer plus précisément les porteurs des allèles à risque soit par l'analyse de neurones issus de cellules pluripotentes induites (IPSC) soit par l'analyse en imagerie cérébrale. Au niveau cellulaire, il serait intéressant de quantifier la quantité de transcrits *CNTNAP2* chez les porteurs et les non porteurs de l'allèle à risque et aussi de voir s'il y a une différence au niveau de la morphologie des neurones, des synapses, des épines dendritiques et des réseaux neuronaux. Une question reste à définir dans quel type de neurones nous devons différencier les IPSC. Au vu des résultats d'expression, nous pourrions commencer par des neurones corticaux comme le laboratoire vient de le faire en collaboration avec Alexandra Benchoua pour l'étude des mutations *SHANK3* (Boissart, Poulet et al. 2013). Et si les résultats sont concluants, nous pourrions utiliser ces cellules pour tester des molécules pour restaurer les fonctions altérées. La question reste ouverte sur la conséquence des variants sur les possibles anomalies de régulation de *CNTNAP2* (plus ou moins de transcrits ?, epissage aberrant ?) Une autre piste est aussi celle des différents facteurs de transcription qui régulent CNTNAP2. L'identification de ces facteurs de transcription et de leurs réseaux de gènes cibles permettra peut être de mieux interpréter les analyses génome entier en ciblant ces réseaux de gènes et de cibler des traitements plus efficacement pour les patients. De plus une analyse de séquençage à grande échelle des allèles à risque définie au cours de nos travaux dans la famille PL, permettrait de définir si cet haplotype à risque est plus représenté chez les patients dyslexiques que chez les témoins, voire à l'élargir à d'autres pathologies comme les TSA et la schizophrénie. Des analyses en imagerie pourraient aussi nous renseigner sur les circuits différemment activés chez les porteurs de l'allèle à risque. De nouveau, les questions restent ouvertes si c'est une hypo- ou une hyper-activation ? et dans quelles régions du cerveau ? Enfin, chez un individu, la mutation CNTNAP2 n'est probablement pas le seul facteur génétique

impliqué. Il serait donc intéressent de réaliser une étude génétique (puce à ADN, exome ou génome entier) des patients (et de leurs parents et germains) porteurs de mutations CNTNAP2 afin d'identifier d'autres mutations. Cette étude minutieuse permettrait de mieux comprendre l'hétérogénéité de ces maladies associées à CNTNAP2.

Perspectives générales

L'évolution a modelé nos capacités cognitives, en développant certaines aires du cerveau grâce à la fonction de certains gènes comme très probablement *FOXP2* et *CNTNAP2* pour le langage. La génétique peut ainsi permettre d'identifier des facteurs qui influencent le cerveau au cours de son développement. Néanmoins, d'autres facteurs sont importants comme les mécanismes épigénétiques sans parler des effets multiples de l'environnement. Il est important de rappeler que le développement du cerveau chez l'homme ne peut être réduit au génome et encore moins à un seul gène.

References

Abrahams, B. S. and D. H. Geschwind (2008). "Advances in autism genetics: on the threshold of a new neurobiology." Nat Rev Genet **9**(5): 341-355.

Abrahams, B. S., D. Tentler, J. V. Perederiy, M. C. Oldham, G. Coppola and D. H. Geschwind (2007). "Genome-wide analyses of human perisylvian cerebral cortical patterning." Proc Natl Acad Sci U S A **104**(45): 17849-17854.

Al-Murrani, A., F. Ashton, S. Aftimos, A. M. George and D. R. Love (2012). "Amino-Terminal Microdeletion within the CNTNAP2 Gene Associated with Variable Expressivity of Speech Delay." Case Rep Genet **2012**: 172408.

Alarcon, M., B. S. Abrahams, J. L. Stone, J. A. Duvall, J. V. Perederiy, J. M. Bomar, J. Sebat, M. Wigler, C. L. Martin, D. H. Ledbetter*, et al.* (2008). "Linkage, association, and gene-expression analyses identify CNTNAP2 as an autism-susceptibility gene." Am J Hum Genet **82**(1): 150-159.

Alkan, C., B. P. Coe and E. E. Eichler (2011). "Genome structural variation discovery and genotyping." Nat Rev Genet **12**(5): 363-376.

Amiet, C., I. Gourfinkel-An, A. Bouzamondo, S. Tordjman, M. Baulac, P. Lechat, L. Mottron and D. Cohen (2008). "Epilepsy in autism is associated with intellectual disability and gender: evidence from a meta-analysis." Biol Psychiatry **64**(7): 577-582.

Amir, R. E., I. B. Van den Veyver, M. Wan, C. Q. Tran, U. Francke and H. Y. Zoghbi (1999). "Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2." Nature Genetics **23**(2): 185-188.

Andari, E., J. R. Duhamel, T. Zalla, E. Herbrecht, M. Leboyer and A. Sirigu (2010). "Promoting social behavior with oxytocin in high-functioning autism spectrum disorders." Proc Natl Acad Sci U S A **107**(9): 4389-4394.

Anney, R., L. Klei, D. Pinto, J. Almeida, E. Bacchelli, G. Baird, N. Bolshakova, S. Bolte, P. F. Bolton, T. Bourgeron*, et al.* (2012). "Individual common variants exert weak effects on the risk for autism spectrum disorderspi." Hum Mol Genet.

Anthoni, H., M. Zucchelli, H. Matsson, B. Muller-Myhsok, I. Fransson, J. Schumacher, S. Massinen, P. Onkamo, A. Warnke, H. Griesemann*, et al.* (2007). "A locus on 2p12 containing the co-regulated MRPL19 and C2ORF3 genes is associated to dyslexia." Hum Mol Genet **16**(6): 667-677.

Arking, D. E., D. J. Cutler, C. W. Brune, T. M. Teslovich, K. West, M. Ikeda, A. Rea, M. Guy, S. Lin, E. H. Cook*, et al.* (2008). "A common genetic variant in the neurexin superfamily member CNTNAP2 increases familial risk of autism." Am J Hum Genet **82**(1): 160-164.

Arons, M. H., C. J. Thynne, A. M. Grabrucker, D. Li, M. Schoen, J. E. Cheyne, T. M. Boeckers, J. M. Montgomery and C. C. Garner (2012). "Autism-Associated Mutations in ProSAP2/Shank3 Impair Synaptic Transmission and Neurexin-Neuroligin-Mediated Transsynaptic Signaling." J Neurosci **32**(43): 14966-14978.

Arroyo, E. J., T. Xu, S. Poliak, M. Watson, E. Peles and S. S. Scherer (2001). "Internodal specializations of myelinated axons in the central nervous system." Cell Tissue Res **305**(1): 53-66.

Ascano, M., Jr., N. Mukherjee, P. Bandaru, J. B. Miller, J. D. Nusbaum, D. L. Corcoran, C. Langlois, M. Munschauer, S. Dewell, M. Hafner*, et al.* (2012). "FMRP targets distinct mRNA sequence elements to regulate protein expression." Nature **492**(7429): 382-386.

Asperger, H. (1944). "Die "autistischen Psychopathen" im Kindesalter." Arch Psychiatr Nervenkr **177**: 76-137.
Auerbach, B. D., E. K. Osterweil and M. F. Bear (2011). "Mutations causing syndromic autism define an axis of synaptic pathophysiology." Nature **480**(7375): 63-68.
Bai, S., K. Ghoshal and S. T. Jacob (2006). "Identification of T-cadherin as a novel target of DNA methyltransferase 3B and its role in the suppression of nerve growth factor-mediated neurite outgrowth in PC12 cells." J Biol Chem **281**(19): 13604-13611.
Bailey, A., A. Le Couteur, I. Gottesman, P. Bolton, E. Simonoff, E. Yuzda and M. Rutter (1995). "Autism as a strongly genetic disorder: evidence from a British twin study." Psychological Medicine **25**(1): 63-77.
Bailey, A., S. Palferman, L. Heavey and A. Le Couteur (1998). "Autism: the phenotype in relatives." Journal of Autism & Developmental Disorders **28**(5): 369-392.
Bailey, D. B., M. Raspa, D. Holiday, E. Bishop and M. Olmsted (2009). "Functional skills of individuals with fragile x syndrome: a lifespan cross-sectional analysis." Am J Intellect Dev Disabil **114**(4): 289-303.
Bakkaloglu, B., B. J. O'Roak, A. Louvi, A. R. Gupta, J. F. Abelson, T. M. Morgan, K. Chawarska, A. Klin, A. G. Ercan-Sencicek, A. A. Stillman*, et al.* (2008). "Molecular cytogenetic analysis and resequencing of contactin associated protein-like 2 in autism spectrum disorders." Am J Hum Genet **82**(1): 165-173.
Ballif, B. C., S. A. Hornor, E. Jenkins, S. Madan-Khetarpal, U. Surti, K. E. Jackson, A. Asamoah, P. L. Brock, G. C. Gowans, R. L. Conway*, et al.* (2007). "Discovery of a previously unrecognized microdeletion syndrome of 16p11.2-p12.2." Nat Genet **39**(9): 1071-1073.
Baron-Cohen, S., E. Ashwin, C. Ashwin, T. Tavassoli and B. Chakrabarti (2009). "Talent in autism: hyper-systemizing, hyper-attention to detail and sensory hypersensitivity." Philos Trans R Soc Lond B Biol Sci **364**(1522): 1377-1383.
Baron-Cohen, S., S. Wheelwright, R. Skinner, J. Martin and E. Clubley (2001). "The autism-spectrum quotient (AQ): evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians." J Autism Dev Disord **31**(1): 5-17.
Barreiro, L. B., G. Laval, H. Quach, E. Patin and L. Quintana-Murci (2008). "Natural selection has driven population differentiation in modern humans." Nat Genet **40**(3): 340-345.
Bates, T. C., P. A. Lind, M. Luciano, G. W. Montgomery, N. G. Martin and M. J. Wright (2010). "Dyslexia and DYX1C1: deficits in reading and spelling associated with a missense mutation." Mol Psychiatry **15**(12): 1190-1196.
Bates, T. C., M. Luciano, A. Castles, M. Coltheart, M. J. Wright and N. G. Martin (2007). "Replication of reported linkages for dyslexia and spelling and suggestive evidence for novel regions on chromosomes 4 and 17." Eur J Hum Genet **15**(2): 194-203.
Battaglia, A., A. Novelli, L. Bernardini, R. Igliozzi and B. Parrini (2009). "Further characterization of the new microdeletion syndrome of 16p11.2-p12.2." Am J Med Genet A **149A**(6): 1200-1204.
Baudouin, S. J., J. Gaudias, S. Gerharz, L. Hatstatt, K. Zhou, P. Punnakkal, K. F. Tanaka, W. Spooren, R. Hen, C. I. De Zeeuw*, et al.* (2012). "Shared synaptic pathophysiology in syndromic and nonsyndromic rodent models of autism." Science **338**(6103): 128-132.
Baumgartner, S., J. T. Littleton, K. Broadie, M. A. Bhat, R. Harbecke, J. A. Lengyel, R. Chiquet-Ehrismann, A. Prokop and H. J. Bellen (1996). "A Drosophila neurexin is

required for septate junction and blood-nerve barrier formation and function." Cell **87**(6): 1059-1068.
Bayarsaihan, D., N. Bitchevaia, B. Enkhmandakh, M. I. Tussie-Luna, J. F. Leckman, A. Roy and F. Ruddle (2003). "Expression of BEN, a member of TFII-I family of transcription factors, during mouse pre- and postimplantation development." Gene Expr Patterns **3**(5): 579-589.
Bellini, G., C. Bravaccio, F. Calamoneri, M. Donatella Cocuzza, P. Fiorillo, A. Gagliano, D. Mazzone, E. M. del Giudice, G. Scuccimarra, R. Militerni*, et al.* (2005). "No evidence for association between dyslexia and DYX1C1 functional variants in a group of children and adolescents from Southern Italy." J Mol Neurosci **27**(3): 311-314.
Belloso, J. M., I. Bache, M. Guitart, M. R. Caballin, C. Halgren, M. Kirchhoff, H. H. Ropers, N. Tommerup and Z. Tumer (2007). "Disruption of the CNTNAP2 gene in a t(7;15) translocation family without symptoms of Gilles de la Tourette syndrome." Eur J Hum Genet **15**(6): 711-713.
Belmonte, M. K., M. Gomot and S. Baron-Cohen (2009). "Visual attention in autism families: 'unaffected' sibs share atypical frontal activation." J Child Psychol Psychiatry.
Ben-Shachar, S., B. Lanpher, J. R. German, M. Qasaymeh, L. Potocki, S. C. Nagamani, L. M. Franco, A. Malphrus, G. W. Bottenfield, J. E. Spence*, et al.* (2009). "Microdeletion 15q13.3: a locus with incomplete penetrance for autism, mental retardation, and psychiatric disorders." J Med Genet **46**(6): 382-388.
Berkel, S., W. Tang, M. Trevino, M. Vogt, H. A. Obenhaus, P. Gass, S. W. Scherer, R. Sprengel, G. Schratt and G. A. Rappold (2011). "Inherited and de novo SHANK2 variants associated with autism spectrum disorder impair neuronal morphogenesis and physiology." Hum Mol Genet **21**(2): 344-357.
Berry-Kravis, E., D. Hessl, S. Coffey, C. Hervey, A. Schneider, J. Yuhas, J. Hutchison, M. Snape, M. Tranfaglia, D. V. Nguyen*, et al.* (2009). "A pilot open label, single dose trial of fenobam in adults with fragile X syndrome." J Med Genet **46**(4): 266-271.
Bijlsma, E. K., A. C. Gijsbers, J. H. Schuurs-Hoeijmakers, A. van Haeringen, D. E. Fransen van de Putte, B. M. Anderlid, J. Lundin, P. Lapunzina, L. A. Perez Jurado, B. Delle Chiaie*, et al.* (2009). "Extending the phenotype of recurrent rearrangements of 16p11.2: deletions in mentally retarded patients without autism and in normal individuals." Eur J Med Genet **52**(2-3): 77-87.
Bishop, D. V. (2003). "Autism and specific language impairment: categorical distinction or continuum?" Novartis Found Symp **251**: 213-226; discussion 226-234, 281-297.
Bishop, R., E. Gobright, V. Nene, S. Morzaria, A. Musoke and B. Sohanpal (2000). "Polymorphic open reading frames encoding secretory proteins are located less than 3 kilobases from Theileria parva telomeres." Mol Biochem Parasitol **110**(2): 359-371.
Bittel, D. C., N. Kibiryeva and M. G. Butler (2006). "Expression of 4 genes between chromosome 15 breakpoints 1 and 2 and behavioral outcomes in Prader-Willi syndrome." Pediatrics **118**(4): e1276-1283.
Boissart, C., A. Poulet, P. Georges, H. Darville, E. Julita, R. Delorme, T. Bourgeron, M. Peschanski and A. Benchoua (2013). "Differentiation from human pluripotent stem cells of cortical neurons of the superficial layers amenable to psychiatric disease modeling and high-throughput drug screening." Transl Psychiatry **3**: e294.
Bourgeron, T. (2009). "A synaptic trek to autism." Curr Opin Neurobiol **19**(2): 231-234.
Brkanac, Z., N. H. Chapman, M. M. Matsushita, L. Chun, K. Nielsen, E. Cochrane, V. W. Berninger, E. M. Wijsman and W. H. Raskind (2007). "Evaluation of candidate

genes for DYX1 and DYX2 in families with dyslexia." Am J Med Genet B Neuropsychiatr Genet **144B**(4): 556-560.
Burbridge, T. J., Y. Wang, A. J. Volz, V. J. Peschansky, L. Lisann, A. M. Galaburda, J. J. Lo Turco and G. D. Rosen (2008). "Postnatal analysis of the effect of embryonic knockdown and overexpression of candidate dyslexia susceptibility gene homolog Dcdc2 in the rat." Neuroscience **152**(3): 723-733.
Burnside, R. D., R. Pasion, F. M. Mikhail, A. J. Carroll, N. H. Robin, E. L. Youngs, I. K. Gadi, E. Keitges, V. L. Jaswaney, P. R. Papenhausen*, et al.* (2011). "Microdeletion/microduplication of proximal 15q11.2 between BP1 and BP2: a susceptibility region for neurological dysfunction including developmental and language delay." Hum Genet **130**(4): 517-528.
Butler, M. G., W. Fischer, N. Kibiryeva and D. C. Bittel (2008). "Array comparative genomic hybridization (aCGH) analysis in Prader-Willi syndrome." Am J Med Genet A **146**(7): 854-860.
Camacho-Garcia, R. J., M. I. Planelles, M. Margalef, M. L. Pecero, R. Martinez-Leal, F. Aguilera, E. Vilella, A. Martinez-Mir and F. G. Scholl (2012). "Mutations affecting synaptic levels of neurexin-1beta in autism and mental retardation." Neurobiol Dis **47**(1): 135-143.
Canitano, R. (2007). "Epilepsy in autism spectrum disorders." Eur Child Adolesc Psychiatry **16**(1): 61-66.
Cardon, L. R., S. D. Smith, D. W. Fulker, W. J. Kimberling, B. F. Pennington and J. C. DeFries (1994). "Quantitative trait locus for reading disability on chromosome 6." Science **266**(5183): 276-279.
Casey, J. P., T. Magalhaes, J. M. Conroy, R. Regan, N. Shah, R. Anney, D. C. Shields, B. S. Abrahams, J. Almeida, E. Bacchelli*, et al.* (2011). "A novel approach of homozygous haplotype sharing identifies candidate genes in autism spectrum disorder." Hum Genet **131**(4): 565-579.
Celestino-Soper, P. B., S. Violante, E. L. Crawford, R. Luo, A. C. Lionel, E. Delaby, G. Cai, B. Sadikovic, K. Lee, C. Lo*, et al.* (2012). "A common X-linked inborn error of carnitine biosynthesis may be a risk factor for nondysmorphic autism." Proc Natl Acad Sci U S A **109**(21): 7974-7981.
Chadman, K. K., S. Gong, M. L. Scattoni, S. E. Boltuck, S. U. Gandhy, N. Heintz and J. N. Crawley (2008). "Minimal Aberrant Behavioral Phenotypes of Neuroligin-3 R451C Knockin Mice." Autism Research **1**(3): 147-158.
Chahrour, M. H., T. W. Yu, E. T. Lim, B. Ataman, M. E. Coulter, R. S. Hill, C. R. Stevens, C. R. Schubert, A. A. S. Collaboration, M. E. Greenberg*, et al.* (2012). "Whole-exome sequencing and homozygosity analysis implicate depolarization-regulated neuronal genes in autism." PLoS Genet **8**(4): e1002635.
Chakrabarti, S. and E. Fombonne (2005). "Pervasive developmental disorders in preschool children: confirmation of high prevalence." Am J Psychiatry **162**(6): 1133-1141.
Chapman, N. H., R. P. Igo, J. B. Thomson, M. Matsushita, Z. Brkanac, T. Holzman, V. W. Berninger, E. M. Wijsman and W. H. Raskind (2004). "Linkage analyses of four regions previously implicated in dyslexia: confirmation of a locus on chromosome 15q." Am J Med Genet B Neuropsychiatr Genet **131B**(1): 67-75.
Chari, S. and I. Dworkin (2013). "The conditional nature of genetic interactions: the consequences of wild-type backgrounds on mutational interactions in a genome-wide modifier screen." PLoS Genet **9**(8): e1003661.

Chih, B., S. K. Afridi, L. Clark and P. Scheiffele (2004). "Disorder-associated mutations lead to functional inactivation of neuroligins." Hum Mol Genet **13**(14): 1471-1477.
Chilian, B., H. Abdollahpour, T. Bierhals, I. Haltrich, G. Fekete, I. Nagel, G. Rosenberger and K. Kutsche (2013). "Dysfunction of SHANK2 and CHRNA7 in a patient with intellectual disability and language impairment supports genetic epistasis of the two loci." Clin Genet **doi: 10.1111/cge.12105**.
Christian, S. L., C. W. Brune, J. Sudi, R. A. Kumar, S. Liu, S. Karamohamed, J. A. Badner, S. Matsui, J. Conroy, D. McQuaid*, et al.* (2008). "Novel submicroscopic chromosomal abnormalities detected in autism spectrum disorder." Biol Psychiatry **63**(12): 1111-1117.
Clemm von Hohenberg, C., M. C. Wigand, M. Kubicki, G. Leicht, I. Giegling, S. Karch, A. M. Hartmann, B. Konte, M. Friedl, T. Ballinger*, et al.* (2013). "CNTNAP2 polymorphisms and structural brain connectivity: A diffusion-tensor imaging study." J Psychiatr Res **47**(10): 1349-1356.
Coleman, M. and C. Gillberg (2012). The autisms, Oxford University Press.
Condro, M. C. and S. A. White (2013). "Distribution of language-related Cntnap2 protein in neural circuits critical for vocal learning." J Comp Neurol.
Constantino, J. N., T. Przybeck, D. Friesen and R. D. Todd (2000). "Reciprocal social behavior in children with and without pervasive developmental disorders." Item Corporate Author Charles A Dana Fdn Work Grp Perv **21**(1): 2-11.
Constantino, J. N. and R. D. Todd (2003). "Autistic traits in the general population: a twin study." Arch Gen Psychiatry **60**(5): 524-530.
Cook, E. H., Jr., V. Lindgren, B. L. Leventhal, R. Courchesne, A. Lincoln, C. Shulman, C. Lord and E. Courchesne (1997). "Autism or atypical autism in maternally but not paternally derived proximal 15q duplication." American Journal of Human Genetics **60**(4): 928-934.
Cooper, G. M., B. P. Coe, S. Girirajan, J. A. Rosenfeld, T. H. Vu, C. Baker, C. Williams, H. Stalker, R. Hamid, V. Hannig*, et al.* (2011). "A copy number variation morbidity map of developmental delay." Nat Genet **43**(9): 838-846.
Cope, N., D. Harold, G. Hill, V. Moskvina, J. Stevenson, P. Holmans, M. J. Owen, M. C. O'Donovan and J. Williams (2005). "Strong evidence that KIAA0319 on chromosome 6p is a susceptibility gene for developmental dyslexia." Am J Hum Genet **76**(4): 581-591.
Cope, N. A., G. Hill, M. van den Bree, D. Harold, V. Moskvina, E. K. Green, M. J. Owen, J. Williams and M. C. O'Donovan (2005). "No support for association between dyslexia susceptibility 1 candidate 1 and developmental dyslexia." Mol Psychiatry **10**(3): 237-238.
Coquelle, F. M., T. Levy, S. Bergmann, S. G. Wolf, D. Bar-El, T. Sapir, Y. Brody, I. Orr, N. Barkai, G. Eichele*, et al.* (2006). "Common and divergent roles for members of the mouse DCX superfamily." Cell Cycle **5**(9): 976-983.
Cornish, K. M., C. S. Kogan, L. Li, J. Turk, S. Jacquemont and R. J. Hagerman (2009). "Lifespan changes in working memory in fragile X premutation males." Brain Cogn **69**(3): 551-558.
Corpas, M., E. Bragin, S. Clayton, P. Bevan and H. V. Firth (2012). "Interpretation of genomic copy number variants using DECIPHER." Curr Protoc Hum Genet **Chapter 8**: Unit 8 14.
Courchesne, E., C. M. Karns, H. R. Davis, R. Ziccardi, R. A. Carper, Z. D. Tigue, H. J. Chisum, P. Moses, K. Pierce, C. Lord*, et al.* (2001). "Unusual brain growth patterns in early life in patients with autistic disorder: an MRI study." Neurology **57**(2): 245-254.

Cross-Disorder Group of the Psychiatric Genomics, C., S. H. Lee, S. Ripke, B. M. Neale, S. V. Faraone, S. M. Purcell, R. H. Perlis, B. J. Mowry, A. Thapar, M. E. Goddard*, et al.* (2013). "Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs." Nat Genet **45**(9): 984-994.

Dahdouh, F., H. Anthoni, I. Tapia-Paez, M. Peyrard-Janvid, G. Schulte-Korne, A. Warnke, H. Remschmidt, A. Ziegler, J. Kere, B. Muller-Myhsok*, et al.* (2009). "Further evidence for DYX1C1 as a susceptibility factor for dyslexia." Psychiatr Genet **19**(2): 59-63.

Darnell, J. C., S. J. Van Driesche, C. Zhang, K. Y. Hung, A. Mele, C. E. Fraser, E. F. Stone, C. Chen, J. J. Fak, S. W. Chi*, et al.* (2011). "FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism." Cell **146**(2): 247-261.

Davies, G., A. Tenesa, A. Payton, J. Yang, S. E. Harris, D. Liewald, X. Ke, S. Le Hellard, A. Christoforou, M. Luciano*, et al.* (2011). "Genome-wide association studies establish that human intelligence is highly heritable and polygenic." Mol Psychiatry **16**(10): 996-1005.

de Kovel, C. G., B. Franke, F. A. Hol, J. J. Lebrec, B. Maassen, H. Brunner, G. W. Padberg, J. Platko and D. Pauls (2008). "Confirmation of dyslexia susceptibility loci on chromosomes 1p and 2p, but not 6p in a Dutch sib-pair collection." Am J Med Genet B Neuropsychiatr Genet **147**(3): 294-300.

de Kovel, C. G., F. A. Hol, J. G. Heister, J. J. Willemen, L. A. Sandkuijl, B. Franke and G. W. Padberg (2004). "Genomewide scan identifies susceptibility locus for dyslexia on Xq27 in an extended Dutch family." J Med Genet **41**(9): 652-657.

Deffenbacher, K. E., J. B. Kenyon, D. M. Hoover, R. K. Olson, B. F. Pennington, J. C. DeFries and S. D. Smith (2004). "Refinement of the 6p21.3 quantitative trait locus influencing dyslexia: linkage and association analyses." Hum Genet **115**(2): 128-138.

Delorme, R., E. Ey, R. Toro, M. Leboyer, C. Gillberg and T. Bourgeron (2013). "Progress toward treatments for synaptic defects in autism." Nat Med **19**(6): 685-694.

Dennis, E. L., N. Jahanshad, J. D. Rudie, J. A. Brown, K. Johnson, K. L. McMahon, G. I. de Zubicaray, G. Montgomery, N. G. Martin, M. J. Wright*, et al.* (2011). "Altered structural brain connectivity in healthy carriers of the autism risk gene, CNTNAP2." Brain Connect **1**(6): 447-459.

Dennis, M. Y., S. Paracchini, T. S. Scerri, L. Prokunina-Olsson, J. C. Knight, R. Wade-Martins, P. Coggill, S. Beck, E. D. Green and A. P. Monaco (2009). "A common variant associated with dyslexia reduces expression of the KIAA0319 gene." PLoS Genet **5**(3): e1000436.

Depienne, C., D. Heron, C. Betancur, B. Benyahia, O. Trouillard, D. Bouteiller, A. Verloes, E. LeGuern, M. Leboyer and A. Brice (2007). "Autism, language delay and mental retardation in a patient with 7q11 duplication." J Med Genet **44**(7): 452-458.

Depienne, C., D. Moreno-De-Luca, D. Heron, D. Bouteiller, A. Gennetier, R. Delorme, P. Chaste, J. P. Siffroi, S. Chantot-Bastaraud, B. Benyahia*, et al.* (2009). "Screening for genomic rearrangements and methylation abnormalities of the 15q11-q13 region in autism spectrum disorders." Biol Psychiatry **66**(4): 349-359.

Derecki, N. C., J. C. Cronk, Z. Lu, E. Xu, S. B. Abbott, P. G. Guyenet and J. Kipnis (2012). "Wild-type microglia arrest pathology in a mouse model of Rett syndrome." Nature **484**(7392): 105-109.

Devlin, B. and S. W. Scherer (2012). "Genetic architecture in autism spectrum disorder." Curr Opin Genet Dev **22**(3): 229-237.

Dibbens, L. M., S. Mullen, I. Helbig, H. C. Mefford, M. A. Bayly, S. Bellows, C. Leu, H. Trucks, T. Obermeier, M. Wittig*, et al.* (2009). "Familial and sporadic 15q13.3 microdeletions in idiopathic generalized epilepsy: precedent for disorders with complex inheritance." Hum Mol Genet **18**(19): 3626-3631.

Dobyns, W. B., C. L. Truwit, M. E. Ross, N. Matsumoto, D. T. Pilz, D. H. Ledbetter, J. G. Gleeson, C. A. Walsh and A. J. Barkovich (1999). "Differences in the gyral pattern distinguish chromosome 17-linked and X-linked lissencephaly." Neurology **53**(2): 270-277.

Duflocq, A., F. Chareyre, M. Giovannini, F. Couraud and M. Davenne (2011). "Characterization of the axon initial segment (AIS) of motor neurons and identification of a para-AIS and a juxtapara-AIS, organized by protein 4.1B." BMC Biol **9**: 66.

Durand, C. M., C. Betancur, T. M. Boeckers, J. Bockmann, P. Chaste, F. Fauchereau, G. Nygren, M. Rastam, I. C. Gillberg, H. Anckarsater*, et al.* (2007). "Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders." Nat Genet **39**(1): 25-27.

Durand, C. M., J. Perroy, F. Loll, D. Perrais, L. Fagni, T. Bourgeron, M. Montcouquiol and N. Sans (2011). "SHANK3 mutations identified in autism lead to modification of dendritic spine morphology via an actin-dependent mechanism." Mol Psychiatry **17**(1): 71-84.

Durbin, R. M., G. R. Abecasis, D. L. Altshuler, A. Auton, L. D. Brooks, R. A. Gibbs, M. E. Hurles and G. A. McVean (2010). "A map of human genome variation from population-scale sequencing." Nature **467**(7319): 1061-1073.

Einheber, S., G. Zanazzi, W. Ching, S. Scherer, T. A. Milner, E. Peles and J. L. Salzer (1997). "The axonal membrane protein Caspr, a homologue of neurexin IV, is a component of the septate-like paranodal junctions that assemble during myelination." J Cell Biol **139**(6): 1495-1506.

Elia, J., X. Gai, H. M. Xie, J. C. Perin, E. Geiger, J. T. Glessner, M. D'Arcy, R. deBerardinis, E. Frackelton, C. Kim*, et al.* (2010). "Rare structural variants found in attention-deficit hyperactivity disorder are preferentially associated with neurodevelopmental genes." Mol Psychiatry **15**(6): 637-646.

Elsabbagh, M., G. Divan, Y. J. Koh, Y. S. Kim, S. Kauchali, C. Marcin, C. Montiel-Nava, V. Patel, C. S. Paula, C. Wang*, et al.* (2012). "Global prevalence of autism and other pervasive developmental disorders." Autism Res **5**(3): 160-179.

Enkhmandakh, B., N. Bitchevaia, F. Ruddle and D. Bayarsaihan (2004). "The early embryonic expression of TFII-I during mouse preimplantation development." Gene Expr Patterns **4**(1): 25-28.

Ey, E., C. S. Leblond and T. Bourgeron (2010). "Behavioral profiles of mouse models for autism spectrum disorders." Autism Res **4**(1): 5-16.

Ey, E., M. Yang, A. M. Katz, L. Woldeyohannes, J. L. Silverman, C. S. Leblond, P. Faure, N. Torquet, A. M. Le Sourd, T. Bourgeron*, et al.* (2012). "Absence of deficits in social behaviors and ultrasonic vocalizations in later generations of mice lacking neuroligin4." Genes Brain Behav.

Fagerheim, T., P. Raeymaekers, F. E. Tonnessen, M. Pedersen, L. Tranebjaerg and H. A. Lubs (1999). "A new gene (DYX3) for dyslexia is located on chromosome 2." J Med Genet **36**(9): 664-669.

Fernandez, B. A., W. Roberts, B. Chung, R. Weksberg, S. Meyn, P. Szatmari, A. M. Joseph-George, S. Mackay, K. Whitten, B. Noble*, et al.* (2009). "Phenotypic Spectrum

Associated with De Novo and Inherited Deletions and Duplications at 16p11.2 in Individuals Ascertained for Diagnosis of Autism Spectrum Disorder." J Med Genet.
Fisher, S. E. and J. C. DeFries (2002). "Developmental dyslexia: genetic dissection of a complex cognitive trait." Nat Rev Neurosci **3**(10): 767-780.
Fisher, S. E., C. Francks, A. J. Marlow, I. L. MacPhie, D. F. Newbury, L. R. Cardon, Y. Ishikawa-Brush, A. J. Richardson, J. B. Talcott, J. Gayan*, et al.* (2002). "Independent genome-wide scans identify a chromosome 18 quantitative-trait locus influencing dyslexia." Nat Genet **30**(1): 86-91.
Fisher, S. E., A. J. Marlow, J. Lamb, E. Maestrini, D. F. Williams, A. J. Richardson, D. E. Weeks, J. F. Stein and A. P. Monaco (1999). "A quantitative-trait locus on chromosome 6p influences different aspects of developmental dyslexia." Am J Hum Genet **64**(1): 146-156.
Fisher, S. E., F. Vargha-Khadem, K. E. Watkins, A. P. Monaco and M. E. Pembrey (1998). "Localisation of a gene implicated in a severe speech and language disorder." Nature Genetics **18**(2): 168-170.
Folstein, S. and M. Rutter (1977). "Infantile autism: a genetic study of 21 twin pairs." Journal of Child Psychology & Psychiatry & Allied Disciplines **18**(4): 297-321.
Fombonne, E. (2003). "Epidemiological surveys of autism and other pervasive developmental disorders: an update." J Autism Dev Disord **33**(4): 365-382.
Forrest, M., R. M. Chapman, A. M. Doyle, C. L. Tinsley, A. Waite and D. J. Blake (2012). "Functional analysis of TCF4 missense mutations that cause Pitt-Hopkins syndrome." Hum Mutat **33**(12): 1676-1686.
Francks, C., S. E. Fisher, R. K. Olson, B. F. Pennington, S. D. Smith, J. C. DeFries and A. P. Monaco (2002). "Fine mapping of the chromosome 2p12-16 dyslexia susceptibility locus: quantitative association analysis and positional candidate genes SEMA4F and OTX1." Psychiatr Genet **12**(1): 35-41.
Francks, C., I. L. MacPhie and A. P. Monaco (2002). "The genetic basis of dyslexia." Lancet Neurol **1**(8): 483-490.
Francks, C., S. Paracchini, S. D. Smith, A. J. Richardson, T. S. Scerri, L. R. Cardon, A. J. Marlow, I. L. MacPhie, J. Walter, B. F. Pennington*, et al.* (2004). "A 77-kilobase region of chromosome 6p22.2 is associated with dyslexia in families from the United Kingdom and from the United States." Am J Hum Genet **75**(6): 1046-1058.
Freitag, C. M. (2007). "The genetics of autistic disorders and its clinical relevance: a review of the literature." Mol Psychiatry **12**(1): 2-22.
Friedman, J. I., T. Vrijenhoek, S. Markx, I. M. Janssen, W. A. van der Vliet, B. H. Faas, N. V. Knoers, W. Cahn, R. S. Kahn, L. Edelmann*, et al.* (2008). "CNTNAP2 gene dosage variation is associated with schizophrenia and epilepsy." Mol Psychiatry **13**(3): 261-266.
Galaburda, A. M. and T. L. Kemper (1979). "Cytoarchitectonic abnormalities in developmental dyslexia: a case study." Ann Neurol **6**(2): 94-100.
Galaburda, A. M., G. F. Sherman, G. D. Rosen, F. Aboitiz and N. Geschwind (1985). "Developmental dyslexia: four consecutive patients with cortical anomalies." Ann Neurol **18**(2): 222-233.
Gauthier, J., N. Champagne, R. G. Lafreniere, L. Xiong, D. Spiegelman, E. Brustein, M. Lapointe, H. Peng, M. Cote, A. Noreau*, et al.* (2010). "De novo mutations in the gene encoding the synaptic scaffolding protein SHANK3 in patients ascertained for schizophrenia." Proc Natl Acad Sci U S A **107**(17): 7863-7868.

Gillberg, C. (2010). "The ESSENCE in child psychiatry: Early Symptomatic Syndromes Eliciting Neurodevelopmental Clinical Examinations." Res Dev Disabil **31**(6): 1543-1551.
Gillberg, C. and H. Soderstrorn (2003). "Learning disability." Lancet **362**(9386): 811-821.
Gilman, S. R., I. Iossifov, D. Levy, M. Ronemus, M. Wigler and D. Vitkup (2011). "Rare de novo variants associated with autism implicate a large functional network of genes involved in formation and function of synapses." Neuron **70**(5): 898-907.
Girirajan, S., J. A. Rosenfeld, B. P. Coe, S. Parikh, N. Friedman, A. Goldstein, R. A. Filipink, J. S. McConnell, B. Angle, W. S. Meschino*, et al.* (2012). "Phenotypic heterogeneity of genomic disorders and rare copy-number variants." N Engl J Med **367**(14): 1321-1331.
Girirajan, S., J. A. Rosenfeld, G. M. Cooper, F. Antonacci, P. Siswara, A. Itsara, L. Vives, T. Walsh, S. E. McCarthy, C. Baker*, et al.* (2010). "A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay." Nat Genet **42**(3): 203-209.
Glessner, J. T., K. Wang, G. Cai, O. Korvatska, C. E. Kim, S. Wood, H. Zhang, A. Estes, C. W. Brune, J. P. Bradfield*, et al.* (2009). "Autism genome-wide copy number variation reveals ubiquitin and neuronal genes." Nature **459**(7246): 569-573.
Golzio, C., J. Willer, M. E. Talkowski, E. C. Oh, Y. Taniguchi, S. Jacquemont, A. Reymond, M. Sun, A. Sawa, J. F. Gusella*, et al.* (2012). "KCTD13 is a major driver of mirrored neuroanatomical phenotypes of the 16p11.2 copy number variant." Nature **485**(7398): 363-367.
Graham, S. A. and S. E. Fisher (2013). "Decoding the genetics of speech and language." Curr Opin Neurobiol **23**(1): 43-51.
Gregor, A., B. Albrecht, I. Bader, E. K. Bijlsma, A. B. Ekici, H. Engels, K. Hackmann, D. Horn, J. Hoyer, J. Klapecki*, et al.* (2011). "Expanding the clinical spectrum associated with defects in CNTNAP2 and NRXN1." BMC Med Genet **12**: 106.
Grigorenko, E. L., F. B. Wood, M. S. Meyer, L. A. Hart, W. C. Speed, A. Shuster and D. L. Pauls (1997). "Susceptibility loci for distinct components of developmental dyslexia on chromosomes 6 and 15." Am J Hum Genet **60**(1): 27-39.
Grigorenko, E. L., F. B. Wood, M. S. Meyer and D. L. Pauls (2000). "Chromosome 6p influences on different dyslexia-related cognitive processes: further confirmation." Am J Hum Genet **66**(2): 715-723.
Grigorenko, E. L., F. B. Wood, M. S. Meyer, J. E. Pauls, L. A. Hart and D. L. Pauls (2001). "Linkage studies suggest a possible locus for developmental dyslexia on chromosome 1p." Am J Med Genet **105**(1): 120-129.
Gross-Tsur, V., O. Manor and R. S. Shalev (1996). "Developmental dyscalculia: prevalence and demographic features." Dev Med Child Neurol **38**(1): 25-33.
Guan, J. S., S. J. Haggarty, E. Giacometti, J. H. Dannenberg, N. Joseph, J. Gao, T. J. Nieland, Y. Zhou, X. Wang, R. Mazitschek*, et al.* (2009). "HDAC2 negatively regulates memory formation and synaptic plasticity." Nature **459**(7243): 55-60.
Guilmatre, A., G. Huguet, R. Delorme and T. Bourgeron (2013). "The emerging role of SHANK genes in neuropsychiatric disorders." Dev Neurobiol.
Guy, J., J. Gan, J. Selfridge, S. Cobb and A. Bird (2007). "Reversal of neurological defects in a mouse model of Rett syndrome." Science **315**(5815): 1143-1147.
Hallmayer, J., S. Cleveland, A. Torres, J. Phillips, B. Cohen, T. Torigoe, J. Miller, A. Fedele, J. Collins, K. Smith*, et al.* (2011). "Genetic heritability and shared environmental factors among twin pairs with autism." Arch Gen Psychiatry **68**(11): 1095-1102.

Hannula-Jouppi, K., N. Kaminen-Ahola, M. Taipale, R. Eklund, J. Nopola-Hemmi, H. Kaariainen and J. Kere (2005). "The axon guidance receptor gene ROBO1 is a candidate gene for developmental dyslexia." PLoS Genet **1**(4): e50.

Hao da, C., Y. Feng, R. Xiao and P. G. Xiao (2011). "Non-neutral nonsynonymous single nucleotide polymorphisms in human ABC transporters: the first comparison of six prediction methods." Pharmacol Rep **63**(4): 924-934.

Harold, D., S. Paracchini, T. Scerri, M. Dennis, N. Cope, G. Hill, V. Moskvina, J. Walter, A. J. Richardson, M. J. Owen*, et al.* (2006). "Further evidence that the KIAA0319 gene confers susceptibility to developmental dyslexia." Mol Psychiatry **11**(12): 1085-1091, 1061.

Hartley, S. L., W. E. Maclean, Jr., M. G. Butler, J. Zarcone and T. Thompson (2005). "Maladaptive behaviors and risk factors among the genetic subtypes of Prader-Willi syndrome." Am J Med Genet A **136**(2): 140-145.

Hartley, S. L., D. M. Sikora and R. McCoy (2008). "Prevalence and risk factors of maladaptive behaviour in young children with Autistic Disorder." J Intellect Disabil Res **52**(10): 819-829.

Helbig, I., H. C. Mefford, A. J. Sharp, M. Guipponi, M. Fichera, A. Franke, H. Muhle, C. de Kovel, C. Baker, S. von Spiczak*, et al.* (2009). "15q13.3 microdeletions increase risk of idiopathic generalized epilepsy." Nat Genet **41**(2): 160-162.

Hiramoto, K., M. Negishi and H. Katoh (2006). "Dock4 is regulated by RhoG and promotes Rac-dependent cell migration." Exp Cell Res **312**(20): 4205-4216.

Horresh, I., V. Bar, J. L. Kissil and E. Peles (2010). "Organization of myelinated axons by Caspr and Caspr2 requires the cytoskeletal adapter protein 4.1B." J Neurosci **30**(7): 2480-2489.

Horresh, I., S. Poliak, S. Grant, D. Bredt, M. N. Rasband and E. Peles (2008). "Multiple molecular interactions determine the clustering of Caspr2 and Kv1 channels in myelinated axons." J Neurosci **28**(52): 14213-14222.

Hsiung, G. Y., B. J. Kaplan, T. L. Petryshen, S. Lu and L. L. Field (2004). "A dyslexia susceptibility locus (DYX7) linked to dopamine D4 receptor (DRD4) region on chromosome 11p15.5." Am J Med Genet B Neuropsychiatr Genet **125B**(1): 112-119.

Huc-Chabrolle, M., C. Charon, A. Guilmatre, P. Vourc'h, G. Tripi, M. A. Barthez, E. Sizaret, R. A. Thepault, S. Le Gallic, J. Hager*, et al.* (2013). "Xq27 FRAXA locus is a strong candidate for dyslexia: evidence from a genome-wide scan in French families." Behav Genet **43**(2): 132-140.

Huguet, G., E. Ey and T. Bourgeron (2013). "The Genetic Landscapes of Autism Spectrum Disorders." Annu Rev Genomics Hum Genet. **14**: 191-213.

Ibrahim, S. H., R. G. Voigt, S. K. Katusic, A. L. Weaver and W. J. Barbaresi (2009). "Incidence of gastrointestinal symptoms in children with autism: a population-based study." Pediatrics **124**(2): 680-686.

Igo, R. P., Jr., N. H. Chapman, V. W. Berninger, M. Matsushita, Z. Brkanac, J. H. Rothstein, T. Holzman, K. Nielsen, W. H. Raskind and E. M. Wijsman (2006). "Genomewide scan for real-word reading subphenotypes of dyslexia: novel chromosome 13 locus and genetic complexity." Am J Med Genet B Neuropsychiatr Genet **141B**(1): 15-27.

International Schizophrenia Consortium (2008). "Rare chromosomal deletions and duplications increase risk of schizophrenia." Nature **455**(7210): 237-241.

Iossifov, I., M. Ronemus, D. Levy, Z. Wang, I. Hakker, J. Rosenbaum, B. Yamrom, Y. H. Lee, G. Narzisi, A. Leotta*, et al.* (2012). "De novo gene disruptions in children on the autistic spectrum." Neuron **74**(2): 285-299.

Itsara, A., H. Wu, J. D. Smith, D. A. Nickerson, I. Romieu, S. J. London and E. E. Eichler (2010). "De novo rates and selection of large copy number variation." Genome Res **20**(11): 1469-1481.

Jackman, C., N. D. Horn, J. P. Molleston and D. K. Sokol (2009). "Gene associated with seizures, autism, and hepatomegaly in an Amish girl." Pediatr Neurol **40**(4): 310-313.

Jacquemont, M. L., D. Sanlaville, R. Redon, O. Raoul, V. Cormier-Daire, S. Lyonnet, J. Amiel, M. Le Merrer, D. Heron, M. C. de Blois*, et al.* (2006). "Array-based comparative genomic hybridisation identifies high frequency of cryptic chromosomal rearrangements in patients with syndromic autism spectrum disorders." J Med Genet **43**(11): 843-849.

Jacquemont, S., A. Curie, V. des Portes, M. G. Torrioli, E. Berry-Kravis, R. J. Hagerman, F. J. Ramos, K. Cornish, Y. He, C. Paulding*, et al.* (2011). "Epigenetic modification of the FMR1 gene in fragile X syndrome is associated with differential response to the mGluR5 antagonist AFQ056." Sci Transl Med **3**(64): 64ra61.

Jacquemont, S., A. Reymond, F. Zufferey, L. Harewood, R. G. Walters, Z. Kutalik, D. Martinet, Y. Shen, A. Valsesia, N. D. Beckmann*, et al.* (2011). "Mirror extreme BMI phenotypes associated with gene dosage at the chromosome 16p11.2 locus." Nature **478**(7367): 97-102.

Jamadar, S., N. R. Powers, S. A. Meda, J. Gelernter, J. R. Gruen and G. D. Pearlson (2011). "Genetic influences of cortical gray matter in language-related regions in healthy controls and schizophrenia." Schizophr Res **129**(2-3): 141-148.

Jamain, S., H. Quach, C. Betancur, M. Rastam, C. Colineaux, I. C. Gillberg, H. Soderstrom, B. Giros, M. Leboyer, C. Gillberg*, et al.* (2003). "Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism." Nat Genet **34**(1): 27-29.

Ji, W., T. Li, Y. Pan, H. Tao, K. Ju, Z. Wen, Y. Fu, Z. An, Q. Zhao, T. Wang*, et al.* (2013). "CNTNAP2 is significantly associated with schizophrenia and major depression in the Han Chinese population." Psychiatry Res **207**(3): 225-228.

K., S. (2001). "T he heritability of language : A review and metaanalysis of twin, adoption, and linkage studies." Language **77**(4): 647-723.

Kadesjo, B. and C. Gillberg (2001). "The comorbidity of ADHD in the general population of Swedish school-age children." J Child Psychol Psychiatry **42**(4): 487-492.

Kalnak, N., M. Peyrard-Janvid, B. Sahlen and H. Forssberg (2012). "Family history interview of a broad phenotype in specific language impairment and matched controls." Genes Brain Behav.

Kaminen, N., K. Hannula-Jouppi, M. Kestila, P. Lahermo, K. Muller, M. Kaaranen, B. Myllyluoma, A. Voutilainen, H. Lyytinen, J. Nopola-Hemmi*, et al.* (2003). "A genome scan for developmental dyslexia confirms linkage to chromosome 2p11 and suggests a new locus on 7q32." J Med Genet **40**(5): 340-345.

Kang, H. J., Y. I. Kawasawa, F. Cheng, Y. Zhu, X. Xu, M. Li, A. M. Sousa, M. Pletikos, K. A. Meyer, G. Sedmak*, et al.* (2011). "Spatio-temporal transcriptome of the human brain." Nature **478**(7370): 483-489.

Kanner, L. (1943). "Autistic disturbances of affective contact." Nerv Child **2**: 217-250.

Kaplan, D. E., J. Gayan, J. Ahn, T. W. Won, D. Pauls, R. K. Olson, J. C. DeFries, F. Wood, B. F. Pennington, G. P. Page*, et al.* (2002). "Evidence for linkage and association with reading disability on 6p21.3-22." Am J Hum Genet **70**(5): 1287-1298.

Kelleher, R. J., 3rd and M. F. Bear (2008). "The autistic neuron: troubled translation?" Cell **135**(3): 401-406.

Kirov, G., D. Gumus, W. Chen, N. Norton, L. Georgieva, M. Sari, M. C. O'Donovan, F. Erdogan, M. J. Owen, H. H. Ropers*, et al.* (2008). "Comparative genome hybridization suggests a role for NRXN1 and APBA2 in schizophrenia." Hum Mol Genet **17**(3): 458-465.

Kishino, T., M. Lalande and J. Wagstaff (1997). "UBE3A/E6-AP mutations cause Angelman syndrome." Nat Genet **15**(1): 70-73.

Klei, L., S. J. Sanders, M. T. Murtha, V. Hus, J. K. Lowe, A. J. Willsey, D. Moreno-De-Luca, T. W. Yu, E. Fombonne, D. Geschwind*, et al.* (2012). "Common genetic variants, acting additively, are a major source of risk for autism." Mol Autism **3**(1): 9.

Kolevzon, A., R. Gross and A. Reichenberg (2007). "Prenatal and perinatal risk factors for autism: a review and integration of findings." Arch Pediatr Adolesc Med **161**(4): 326-333.

Kong, A., M. L. Frigge, G. Masson, S. Besenbacher, P. Sulem, G. Magnusson, S. A. Gudjonsson, A. Sigurdsson, A. Jonasdottir, W. S. Wong*, et al.* (2012). "Rate of de novo mutations and the importance of father's age to disease risk." Nature **488**(7412): 471-475.

Konyukh, M., R. Delorme, P. Chaste, C. Leblond, N. Lemiere, G. Nygren, H. Anckarsater, M. Rastam, O. Stahlberg, F. Amsellem*, et al.* (2011). "Variations of the candidate SEZ6L2 gene on Chromosome 16p11.2 in patients with autism spectrum disorders and in human populations." PLoS One **6**(3): e17289.

Kos, M., D. van den Brink, T. M. Snijders, M. Rijpkema, B. Franke, G. Fernandez and P. Hagoort (2012). "CNTNAP2 and language processing in healthy individuals as measured with ERPs." PLoS One **7**(10): e46995.

Kumar, R. A., S. KaraMohamed, J. Sudi, D. F. Conrad, C. Brune, J. A. Badner, T. C. Gilliam, N. J. Nowak, E. H. Cook, Jr., W. B. Dobyns*, et al.* (2008). "Recurrent 16p11.2 microdeletions in autism." Hum Mol Genet **17**(4): 628-638.

Kumar, R. A., C. R. Marshall, J. A. Badner, T. D. Babatz, Z. Mukamel, K. A. Aldinger, J. Sudi, C. W. Brune, G. Goh, S. Karamohamed*, et al.* (2009). "Association and mutation analyses of 16p11.2 autism candidate genes." PLoS One **4**(2): e4582.

Lai, C. S., S. E. Fisher, J. A. Hurst, F. Vargha-Khadem and A. P. Monaco (2001). "A forkhead-domain gene is mutated in a severe speech and language disorder." Nature **413**(6855): 519-523.

Laumonnier, F., F. Bonnet-Brilhault, M. Gomot, R. Blanc, A. David, M. P. Moizard, M. Raynaud, N. Ronce, E. Lemonnier, P. Calvas*, et al.* (2004). "X-linked mental retardation and autism are associated with a mutation in the NLGN4 gene, a member of the neuroligin family." Am J Hum Genet **74**(3): 552-557.

Leblond, C. S., J. Heinrich, R. Delorme, C. Proepper, C. Betancur, G. Huguet, M. Konyukh, P. Chaste, E. Ey, M. Rastam*, et al.* (2012). "Genetic and functional analyses of SHANK2 mutations suggest a multiple hit model of autism spectrum disorders." PLoS Genet **8**(2): e1002521.

Lemonnier, E. and Y. Ben-Ari (2010). "The diuretic bumetanide decreases autistic behaviour in five infants treated during 3 months with no side effects." Acta Paediatr **99**(12): 1885-1888.

Leyfer, O. T., S. E. Folstein, S. Bacalman, N. O. Davis, E. Dinh, J. Morgan, H. Tager-Flusberg and J. E. Lainhart (2006). "Comorbid psychiatric disorders in children with autism: interview development and rates of disorders." J Autism Dev Disord **36**(7): 849-861.

Li, X., Z. Hu, Y. He, Z. Xiong, Z. Long, Y. Peng, F. Bu, J. Ling, G. Xun, X. Mo*, et al.* (2010). "Association analysis of CNTNAP2 polymorphisms with autism in the Chinese Han population." Psychiatr Genet **20**(3): 113-117.

Lichtenstein, P., E. Carlstrom, M. Rastam, C. Gillberg and H. Anckarsater (2010). "The Genetics of Autism Spectrum Disorders and Related Neuropsychiatric Disorders in Childhood." Am J Psychiatry **AiA**: 1–7.

Lim, E. T., S. Raychaudhuri, S. J. Sanders, C. Stevens, A. Sabo, D. G. MacArthur, B. M. Neale, A. Kirby, D. M. Ruderfer, M. Fromer*, et al.* (2013). "Rare complete knockouts in humans: population distribution and significant role in autism spectrum disorders." Neuron **77**(2): 235-242.

Lind, P. A., M. Luciano, M. J. Wright, G. W. Montgomery, N. G. Martin and T. C. Bates (2010). "Dyslexia and DCDC2: normal variation in reading and spelling is associated with DCDC2 polymorphisms in an Australian population sample." Eur J Hum Genet **18**(6): 668-673.

Lindgren, K. A., S. E. Folstein, J. B. Tomblin and H. Tager-Flusberg (2009). "Language and reading abilities of children with autism spectrum disorders and specific language impairment and their first-degree relatives." Autism Res **2**(1): 22-38.

Lindgren, S. D., E. De Renzi and L. C. Richman (1985). "Cross-national comparisons of developmental dyslexia in Italy and the United States." Child Dev **56**(6): 1404-1417.

Lord, C., E. Petkova, V. Hus, W. Gan, F. Lu, D. M. Martin, O. Ousley, L. Guy, R. Bernier, J. Gerdts*, et al.* (2012). "A multisite study of the clinical diagnosis of different autism spectrum disorders." Arch Gen Psychiatry **69**(3): 306-313.

Lord, C., S. Risi, L. Lambrecht, E. H. Cook, Jr., B. L. Leventhal, P. C. DiLavore, A. Pickles and M. Rutter (2000). "The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism." J. Autism Dev. Disord. **30**(3): 205-223.

Lord, C., M. Rutter and A. Le Couteur (1994). "Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders." J Autism Dev Disord **24**(5): 659-685.

Ludwig, K. U., J. Schumacher, G. Schulte-Korne, I. R. Konig, A. Warnke, E. Plume, H. Anthoni, M. Peyrard-Janvid, H. Meng, A. Ziegler*, et al.* (2008). "Investigation of the DCDC2 intron 2 deletion/compound short tandem repeat polymorphism in a large German dyslexia sample." Psychiatr Genet **18**(6): 310-312.

Lundstrom, S., Z. Chang, M. Rastam, C. Gillberg, H. Larsson, H. Anckarsater and P. Lichtenstein (2012). "Autism spectrum disorders and autistic like traits: similar etiology in the extreme end and the normal variation." Arch Gen Psychiatry **69**(1): 46-52.

Lyon GR, S. S., Shaywitz BA (2003). "A definition of dyslexia." Annals of Dyslexia. **53**: 1-14.

M., B. (1993). "Word Recognition and Component Phonological Processing Skills of Adults with Childhood Diagnosis of Dyslexia." Developmental Review **13**(3): 258-268.

Mabb, A. M., M. C. Judson, M. J. Zylka and B. D. Philpot (2011). "Angelman syndrome: insights into genomic imprinting and neurodevelopmental phenotypes." Trends Neurosci **34**(6): 293-303.

Manolio, T. A., F. S. Collins, N. J. Cox, D. B. Goldstein, L. A. Hindorff, D. J. Hunter, M. I. McCarthy, E. M. Ramos, L. R. Cardon, A. Chakravarti*, et al.* (2009). "Finding the missing heritability of complex diseases." Nature **461**(7265): 747-753.

Marino, C., A. Citterio, R. Giorda, A. Facoetti, G. Menozzi, L. Vanzin, M. L. Lorusso, M. Nobile and M. Molteni (2007). "Association of short-term memory with a variant within DYX1C1 in developmental dyslexia." Genes Brain Behav **6**(7): 640-646.

Marino, C., R. Giorda, L. Vanzin, M. Nobile, M. L. Lorusso, C. Baschirotto, L. Riva, M. Molteni and M. Battaglia (2004). "A locus on 15q15-15qter influences dyslexia: further support from a transmission/disequilibrium study in an Italian speaking population." J Med Genet **41**(1): 42-46.

Marshall, C. R., A. Noor, J. B. Vincent, A. C. Lionel, L. Feuk, J. Skaug, M. Shago, R. Moessner, D. Pinto, Y. Ren*, et al.* (2008). "Structural variation of chromosomes in autism spectrum disorder." Am J Hum Genet **82**(2): 477-488.

Matsuura, T., J. S. Sutcliffe, P. Fang, R. J. Galjaard, Y. H. Jiang, C. S. Benton, J. M. Rommens and A. L. Beaudet (1997). "De novo truncating mutations in E6-AP ubiquitin-protein ligase gene (UBE3A) in Angelman syndrome." Nat Genet **15**(1): 74-77.

McArthur, G. M., J. H. Hogben, V. T. Edwards, S. M. Heath and E. D. Mengler (2000). "On the "specifics" of specific reading disability and specific language impairment." J Child Psychol Psychiatry **41**(7): 869-874.

McPheeters, M. L., Z. Warren, N. Sathe, J. L. Bruzek, S. Krishnaswami, R. N. Jerome and J. Veenstra-Vanderweele (2011). "A systematic review of medical treatments for children with autism spectrum disorders." Pediatrics **127**(5): e1312-1321.

Meda, S. A., J. Gelernter, J. R. Gruen, V. D. Calhoun, H. Meng, N. A. Cope and G. D. Pearlson (2008). "Polymorphism of DCDC2 Reveals Differences in Cortical Morphology of Healthy Individuals-A Preliminary Voxel Based Morphometry Study." Brain Imaging Behav **2**(1): 21-26.

Mefford, H. C., H. Muhle, P. Ostertag, S. von Spiczak, K. Buysse, C. Baker, A. Franke, A. Malafosse, P. Genton, P. Thomas*, et al.* (2010). "Genome-wide copy number variation in epilepsy: novel susceptibility loci in idiopathic generalized and focal epilepsies." PLoS Genet **6**(5): e1000962.

Menegoz, M., P. Gaspar, M. Le Bert, T. Galvez, F. Burgaya, C. Palfrey, P. Ezan, F. Arnos and J. A. Girault (1997). "Paranodin, a glycoprotein of neuronal paranodal membranes." Neuron **19**(2): 319-331.

Meng, H., K. Hager, M. Held, G. P. Page, R. K. Olson, B. F. Pennington, J. C. DeFries, S. D. Smith and J. R. Gruen (2005). "TDT-association analysis of EKN1 and dyslexia in a Colorado twin cohort." Hum Genet **118**(1): 87-90.

Meng, H., S. D. Smith, K. Hager, M. Held, J. Liu, R. K. Olson, B. F. Pennington, J. C. DeFries, J. Gelernter, T. O'Reilly-Pol*, et al.* (2005). "DCDC2 is associated with reading disability and modulates neuronal development in the brain." Proc Natl Acad Sci U S A **102**(47): 17053-17058.

Mikhail, F. M., E. J. Lose, N. H. Robin, M. D. Descartes, K. D. Rutledge, S. L. Rutledge, B. R. Korf and A. J. Carroll (2011). "Clinically relevant single gene or intragenic deletions encompassing critical neurodevelopmental genes in patients with developmental delay, mental retardation, and/or autism spectrum disorders." Am J Med Genet A **155A**(10): 2386-2396.

Miles, J. H., T. N. Takahashi, S. Bagby, P. K. Sahota, D. F. Vaslow, C. H. Wang, R. E. Hillman and J. E. Farmer (2005). "Essential versus complex autism: definition of fundamental prognostic subtypes." Am J Med Genet A **135**(2): 171-180.

Miller, D. T., Y. Shen, L. A. Weiss, J. Korn, I. Anselm, C. Bridgemohan, G. F. Cox, H. Dickinson, J. Gentile, D. J. Harris*, et al.* (2009). "Microdeletion/duplication at 15q13.2q13.3 among individuals with features of autism and other neuropsychiatric disorders." J Med Genet **46**(4): 242-248.

Ming, X., M. Brimacombe and G. C. Wagner (2007). "Prevalence of motor impairment in autism spectrum disorders." Brain Dev **29**(9): 565-570.
Moreno-De-Luca, A., S. M. Myers, T. D. Challman, D. Moreno-De-Luca, D. W. Evans and D. H. Ledbetter (2013). "Developmental brain dysfunction: revival and expansion of old concepts based on new genetic evidence." Lancet Neurol **12**(4): 406-414.
Morris, D. W., L. Robinson, D. Turic, M. Duke, V. Webb, C. Milham, E. Hopkin, K. Pound, S. Fernando, M. Easton*, et al.* (2000). "Family-based association mapping provides evidence for a gene for reading disability on chromosome 15q." Hum Mol Genet **9**(5): 843-848.
Morrow, E. M., S. Y. Yoo, S. W. Flavell, T. K. Kim, Y. Lin, R. S. Hill, N. M. Mukaddes, S. Balkhy, G. Gascon, A. Hashmi*, et al.* (2008). "Identifying autism loci and genes by tracing recent shared ancestry." Science **321**(5886): 218-223.
Nakabayashi, K. and S. W. Scherer (2001). "The human contactin-associated protein-like 2 gene (CNTNAP2) spans over 2 Mb of DNA at chromosome 7q35." Genomics **73**(1): 108-112.
Napoli, I., V. Mercaldo, P. P. Boyl, B. Eleuteri, F. Zalfa, S. De Rubeis, D. Di Marino, E. Mohr, M. Massimi, M. Falconi*, et al.* (2008). "The fragile X syndrome protein represses activity-dependent translation through CYFIP1, a new 4E-BP." Cell **134**(6): 1042-1054.
Nava, C., F. Lamari, D. Heron, C. Mignot, A. Rastetter, B. Keren, D. Cohen, A. Faudet, D. Bouteiller, M. Gilleron*, et al.* (2012). "Analysis of the chromosome X exome in patients with autism spectrum disorders identified novel candidate genes, including TMLHE." Transl Psychiatry **2**: e179.
Neale, B. M., Y. Kou, L. Liu, A. Ma'ayan, K. E. Samocha, A. Sabo, C. F. Lin, C. Stevens, L. S. Wang, V. Makarov*, et al.* (2012). "Patterns and rates of exonic de novo mutations in autism spectrum disorders." Nature **485**(7397): 242-245.
Neves-Pereira, M., B. Muller, D. Massie, J. H. Williams, P. C. O'Brien, A. Hughes, S. B. Shen, D. S. Clair and Z. Miedzybrodzka (2009). "Deregulation of EIF4E: a novel mechanism for autism." J Med Genet **46**(11): 759-765.
Newbury, D. F. and A. P. Monaco (2010). "Genetic advances in the study of speech and language disorders." Neuron **68**(2): 309-320.
Newbury, D. F., S. Paracchini, T. S. Scerri, L. Winchester, L. Addis, A. J. Richardson, J. Walter, J. F. Stein, J. B. Talcott and A. P. Monaco (2011). "Investigation of dyslexia and SLI risk variants in reading- and language-impaired subjects." Behav Genet **41**(1): 90-104.
Nopola-Hemmi, J., B. Myllyluoma, T. Haltia, M. Taipale, V. Ollikainen, T. Ahonen, A. Voutilainen, J. Kere and E. Widen (2001). "A dominant gene for developmental dyslexia on chromosome 3." J Med Genet **38**(10): 658-664.
Nopola-Hemmi, J., M. Taipale, T. Haltia, A. E. Lehesjoki, A. Voutilainen and J. Kere (2000). "Two translocations of chromosome 15q associated with dyslexia." J Med Genet **37**(10): 771-775.
O'Roak, B. J., P. Deriziotis, C. Lee, L. Vives, J. J. Schwartz, S. Girirajan, E. Karakoc, A. P. Mackenzie, S. B. Ng, C. Baker*, et al.* (2011). "Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations." Nat Genet **43**(6): 585-589.
O'Roak, B. J., L. Vives, W. Fu, J. D. Egertson, I. B. Stanaway, I. G. Phelps, G. Carvill, A. Kumar, C. Lee, K. Ankenman*, et al.* (2012). "Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders." Science **338**(6114): 1619-1622.

O'Roak, B. J., L. Vives, S. Girirajan, E. Karakoc, N. Krumm, B. P. Coe, R. Levy, A. Ko, C. Lee, J. D. Smith*, et al.* (2012). "Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations." Nature **485**(7397): 246-250.

Ozonoff, S., G. S. Young, A. Carter, D. Messinger, N. Yirmiya, L. Zwaigenbaum, S. Bryson, L. J. Carver, J. N. Constantino, K. Dobkins*, et al.* (2011). "Recurrence risk for autism spectrum disorders: a Baby Siblings Research Consortium study." Pediatrics **128**(3): e488-495.

P., S.-C. L. C. (2003). "Lecture et dyslexie.".

Pagnamenta, A. T., E. Bacchelli, M. V. de Jonge, G. Mirza, T. S. Scerri, F. Minopoli, A. Chiocchetti, K. U. Ludwig, P. Hoffmann, S. Paracchini*, et al.* (2010). "Characterization of a family with rare deletions in CNTNAP5 and DOCK4 suggests novel risk loci for autism and dyslexia." Biol Psychiatry **68**(4): 320-328.

Pagnamenta, A. T., K. Wing, E. S. Akha, S. J. Knight, S. Bolte, G. Schmotzer, E. Duketis, F. Poustka, S. M. Klauck, A. Poustka*, et al.* (2009). "A 15q13.3 microdeletion segregating with autism." Eur J Hum Genet **17**(5): 687-692.

Panaitof, S. C., B. S. Abrahams, H. Dong, D. H. Geschwind and S. A. White (2010). "Language-related Cntnap2 gene is differentially expressed in sexually dimorphic song nuclei essential for vocal learning in songbirds." J Comp Neurol **518**(11): 1995-2018.

Paracchini, S., A. Thomas, S. Castro, C. Lai, M. Paramasivam, Y. Wang, B. J. Keating, J. M. Taylor, D. F. Hacking, T. Scerri*, et al.* (2006). "The chromosome 6p22 haplotype associated with dyslexia reduces the expression of KIAA0319, a novel gene involved in neuronal migration." Hum Mol Genet **15**(10): 1659-1666.

Paulesu, E., J. F. Demonet, F. Fazio, E. McCrory, V. Chanoine, N. Brunswick, S. F. Cappa, G. Cossu, M. Habib, C. D. Frith*, et al.* (2001). "Dyslexia: cultural diversity and biological unity." Science **291**(5511): 2165-2167.

Peca, J., C. Feliciano, J. T. Ting, W. Wang, M. F. Wells, T. N. Venkatraman, C. D. Lascola, Z. Fu and G. Feng (2011). "Shank3 mutant mice display autistic-like behaviours and striatal dysfunction." Nature **472**(7344): 437-442.

Peles, E., M. Nativ, M. Lustig, M. Grumet, J. Schilling, R. Martinez, G. D. Plowman and J. Schlessinger (1997). "Identification of a novel contactin-associated transmembrane receptor with multiple domains implicated in protein-protein interactions." Embo J **16**(5): 978-988.

Penagarikano, O., B. S. Abrahams, E. I. Herman, K. D. Winden, A. Gdalyahu, H. Dong, L. I. Sonnenblick, R. Gruver, J. Almajano, A. Bragin*, et al.* (2011). "Absence of CNTNAP2 leads to epilepsy, neuronal migration abnormalities, and core autism-related deficits." Cell **147**(1): 235-246.

Penagarikano, O., B. S. Abrahams, E. I. Herman, K. D. Winden, A. Gdalyahu, H. M. Dong, L. I. Sonnenblick, R. Gruver, J. Almajano, A. Bragin*, et al.* (2011). "Absence of CNTNAP2 Leads to Epilepsy, Neuronal Migration Abnormalities, and Core Autism-Related Deficits." Cell **147**(1): 235-246.

Penagarikano, O. and D. H. Geschwind (2012). "What does CNTNAP2 reveal about autism spectrum disorder?" Trends Mol Med **18**(3): 156-163.

Peschansky, V. J., T. J. Burbridge, A. J. Volz, C. Fiondella, Z. Wissner-Gross, A. M. Galaburda, J. J. Lo Turco and G. D. Rosen (2010). "The effect of variation in expression of the candidate dyslexia susceptibility gene homolog Kiaa0319 on neuronal migration and dendritic morphology in the rat." Cereb Cortex **20**(4): 884-897.

Peter, B., W. H. Raskind, M. Matsushita, M. Lisowski, T. Vu, V. W. Berninger, E. M. Wijsman and Z. Brkanac (2011). "Replication of CNTNAP2 association with nonword repetition and support for FOXP2 association with timed reading and motor activities in a dyslexia family sample." J Neurodev Disord **3**(1): 39-49.

Petrin, A. L., C. M. Giacheti, L. P. Maximino, D. V. Abramides, S. Zanchetta, N. F. Rossi, A. Richieri-Costa and J. C. Murray (2010). "Identification of a microdeletion at the 7q33-q35 disrupting the CNTNAP2 gene in a Brazilian stuttering case." Am J Med Genet A **152A**(12): 3164-3172.

Petryshen, T. L., B. J. Kaplan, M. Fu Liu, N. S. de French, R. Tobias, M. L. Hughes and L. L. Field (2001). "Evidence for a susceptibility locus on chromosome 6q influencing phonological coding dyslexia." Am J Med Genet **105**(6): 507-517.

Petryshen, T. L., B. J. Kaplan, M. L. Hughes, J. Tzenova and L. L. Field (2002). "Supportive evidence for the DYX3 dyslexia susceptibility gene in Canadian families." J Med Genet **39**(2): 125-126.

Peyrard-Janvid, M., H. Anthoni, P. Onkamo, P. Lahermo, M. Zucchelli, N. Kaminen, K. Hannula-Jouppi, J. Nopola-Hemmi, A. Voutilainen, H. Lyytinen*, et al.* (2004). "Fine mapping of the 2p11 dyslexia locus and exclusion of TACR1 as a candidate gene." Hum Genet **114**(5): 510-516.

Pieretti, M., F. P. Zhang, Y. H. Fu, S. T. Warren, B. A. Oostra, C. T. Caskey and D. L. Nelson (1991). "Absence of expression of the FMR-1 gene in fragile X syndrome." Cell **66**(4): 817-822.

Pinel, P., F. Fauchereau, A. Moreno, A. Barbot, M. Lathrop, D. Zelenika, D. Le Bihan, J. B. Poline, T. Bourgeron and S. Dehaene (2012). "Genetic variants of FOXP2 and KIAA0319/TTRAP/THEM2 locus are associated with altered brain activation in distinct language-related regions." J Neurosci **32**(3): 817-825.

Pinto, D., A. T. Pagnamenta, L. Klei, R. Anney, D. Merico, R. Regan, J. Conroy, T. R. Magalhaes, C. Correia, B. S. Abrahams*, et al.* (2010). "Functional impact of global rare copy number variation in autism spectrum disorders." Nature **466**(7304): 368-372

Platko, J. V., F. B. Wood, I. Pelser, M. Meyer, G. S. Gericke, J. O'Rourke, J. Birns, S. Purcell and D. L. Pauls (2008). "Association of reading disability on chromosome 6p22 in the Afrikaner population." Am J Med Genet B Neuropsychiatr Genet **147B**(7): 1278-1287.

Plomin R., O. M. J. M. n. P. (1994). "The genetic basis of complex human behaviors." Science **264**(5166): 1733-1739.

Poelmans, G., J. K. Buitelaar, D. L. Pauls and B. Franke (2011). "A theoretical molecular network for dyslexia: integrating available genetic findings." Mol Psychiatry **16**(4): 365-382.

Poelmans, G., J. J. Engelen, J. Van Lent-Albrechts, H. J. Smeets, E. Schoenmakers, B. Franke, J. K. Buitelaar, M. Wuisman-Frerker, W. Erens, J. Steyaert*, et al.* (2009). "Identification of novel dyslexia candidate genes through the analysis of a chromosomal deletion." Am J Med Genet B Neuropsychiatr Genet **150B**(1): 140-147.

Poliak, S., L. Gollan, R. Martinez, A. Custer, S. Einheber, J. L. Salzer, J. S. Trimmer, P. Shrager and E. Peles (1999). "Caspr2, a new member of the neurexin superfamily, is localized at the juxtaparanodes of myelinated axons and associates with K+ channels." Neuron **24**(4): 1037-1047.

Poliak, S., D. Salomon, H. Elhanany, H. Sabanay, B. Kiernan, L. Pevny, C. L. Stewart, X. Xu, S. Y. Chiu, P. Shrager*, et al.* (2003). "Juxtaparanodal clustering of Shaker-like K+

channels in myelinated axons depends on Caspr2 and TAG-1." J Cell Biol **162**(6): 1149-1160.
Poot, M., V. Beyer, I. Schwaab, N. Damatova, R. Van't Slot, J. Prothero, S. E. Holder and T. Haaf (2010). "Disruption of CNTNAP2 and additional structural genome changes in a boy with speech delay and autism spectrum disorder." Neurogenetics **11**(1): 81-89.
Rabin, M., X. L. Wen, M. Hepburn, H. A. Lubs, E. Feldman and R. Duara (1993). "Suggestive linkage of developmental dyslexia to chromosome 1p34-p36." Lancet **342**(8864): 178.
Ramocki, M. B. and H. Y. Zoghbi (2008). "Failure of neuronal homeostasis results in common neuropsychiatric phenotypes." Nature **455**(7215): 912-918.
Ramus, F. (2004). "Neurobiology of dyslexia: a reinterpretation of the data." Trends Neurosci **27**(12): 720-726.
Rasband, M. N., E. W. Park, D. Zhen, M. I. Arbuckle, S. Poliak, E. Peles, S. G. Grant and J. S. Trimmer (2002). "Clustering of neuronal potassium channels is independent of their interaction with PSD-95." J Cell Biol **159**(4): 663-672.
Raskind, W. H., R. P. Igo, N. H. Chapman, V. W. Berninger, J. B. Thomson, M. Matsushita, Z. Brkanac, T. Holzman, M. Brown and E. M. Wijsman (2005). "A genome scan in multigenerational families with dyslexia: Identification of a novel locus on chromosome 2q that contributes to phonological decoding efficiency." Mol Psychiatry **10**(7): 699-711.
Redcay, E. and E. Courchesne (2005). "When is the brain enlarged in autism? A meta-analysis of all brain size reports." Biol Psychiatry **58**(1): 1-9.
Reiner, O., F. M. Coquelle, B. Peter, T. Levy, A. Kaplan, T. Sapir, I. Orr, N. Barkai, G. Eichele and S. Bergmann (2006). "The evolving doublecortin (DCX) superfamily." BMC Genomics **7**: 188.
Rodenas-Cuadrado, P., J. Ho and S. C. Vernes (2013). "Shining a light on CNTNAP2: complex functions to complex disorders." Eur J Hum Genet.
Ronald, A., H. Larsson, H. Anckarsater and P. Lichtenstein (2010). "A twin study of autism symptoms in Sweden." Mol Psychiatry **16**(10): 1039-1047.
Rosen, G. D., J. Bai, Y. Wang, C. G. Fiondella, S. W. Threlkeld, J. J. LoTurco and A. M. Galaburda (2007). "Disruption of neuronal migration by RNAi of Dyx1c1 results in neocortical and hippocampal malformations." Cereb Cortex **17**(11): 2562-2572.
Rosenfeld, J. A., K. Leppig, B. C. Ballif, H. Thiese, C. Erdie-Lalena, E. Bawle, S. Sastry, J. E. Spence, A. Bandholz, U. Surti*, et al.* (2009). "Genotype-phenotype analysis of TCF4 mutations causing Pitt-Hopkins syndrome shows increased seizure activity with missense mutations." Genet Med **11**(11): 797-805.
Rosenhall, U., V. Nordin, M. Sandstrom, G. Ahlsen and C. Gillberg (1999). "Autism and hearing loss." J Autism Dev Disord **29**(5): 349-357.
Rossi, E., A. P. Verri, M. G. Patricelli, V. Destefani, I. Ricca, A. Vetro, R. Ciccone, R. Giorda, D. Toniolo, P. Maraschio*, et al.* (2008). "A 12Mb deletion at 7q33-q35 associated with autism spectrum disorders and primary amenorrhea." Eur J Med Genet **51**(6): 631-638.
Rothermundt, M., M. Peters, J. H. Prehn and V. Arolt (2003). "S100B in brain damage and neurodegeneration." Microsc Res Tech **60**(6): 614-632.
Rujescu, D., A. Ingason, S. Cichon, O. P. Pietilainen, M. R. Barnes, T. Toulopoulou, M. Picchioni, E. Vassos, U. Ettinger, E. Bramon*, et al.* (2009). "Disruption of the neurexin 1 gene is associated with schizophrenia." Hum Mol Genet **18**(5): 988-996.

Rutter, M., A. Caspi, D. Fergusson, L. J. Horwood, R. Goodman, B. Maughan, T. E. Moffitt, H. Meltzer and J. Carroll (2004). "Sex differences in developmental reading disability: new findings from 4 epidemiological studies." JAMA **291**(16): 2007-2012.

Sampath, S., S. Bhat, S. Gupta, A. O'Connor, A. B. West, D. E. Arking and A. Chakravarti (2013). "Defining the Contribution of CNTNAP2 to Autism Susceptibility." PLoS One **8**(10): e77906.

Sanders, S. J., A. G. Ercan-Sencicek, V. Hus, R. Luo, M. T. Murtha, D. Moreno-De-Luca, S. H. Chu, M. P. Moreau, A. R. Gupta, S. A. Thomson*, et al.* (2011). "Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism." Neuron **70**(5): 863-885.

Sanders, S. J., M. T. Murtha, A. R. Gupta, J. D. Murdoch, M. J. Raubeson, A. J. Willsey, A. G. Ercan-Sencicek, N. M. Dilullo, N. N. Parikshak, J. L. Stein*, et al.* (2012). "De novo mutations revealed by whole-exome sequencing are strongly associated with autism." Nature **485**: 237-241.

Sato, D., A. C. Lionel, C. S. Leblond, A. Prasad, D. Pinto, S. Walker, I. O'Connor, C. Russell, I. E. Drmic, F. F. Hamdan*, et al.* (2012). "SHANK1 Deletions in Males with Autism Spectrum Disorder." Am J Hum Genet **90**(5): 879-887.

Scerri, T. S., S. E. Fisher, C. Francks, I. L. MacPhie, S. Paracchini, A. J. Richardson, J. F. Stein and A. P. Monaco (2004). "Putative functional alleles of DYX1C1 are not associated with dyslexia susceptibility in a large sample of sibling pairs from the UK." J Med Genet **41**(11): 853-857.

Scerri, T. S. and G. Schulte-Korne (2010). "Genetics of developmental dyslexia." Eur Child Adolesc Psychiatry **19**(3): 179-197.

Scheffer, I. E., B. E. Grinton, S. E. Heron, S. Kivity, Z. Afawi, X. Iona, H. Goldberg-Stern, M. Kinali, I. Andrews, R. Guerrini*, et al.* (2012). "PRRT2 phenotypic spectrum includes sporadic and fever-related infantile seizures." Neurology **79**(21): 2104-2108.

Schenck, A., B. Bardoni, C. Langmann, N. Harden, J. L. Mandel and A. Giangrande (2003). "CYFIP/Sra-1 controls neuronal connectivity in Drosophila and links the Rac1 GTPase pathway to the fragile X protein." Neuron **38**(6): 887-898.

Schmeisser, M. J., E. Ey, S. Wegener, J. Bockmann, A. V. Stempel, A. Kuebler, A. L. Janssen, P. T. Udvardi, E. Shiban, C. Spilker*, et al.* (2012). "Autistic-like behaviours and hyperactivity in mice lacking ProSAP1/Shank2." Nature **486**(7402): 256-260.

Schulte-Korne, G., T. Grimm, M. M. Nothen, B. Muller-Myhsok, S. Cichon, I. R. Vogt, P. Propping and H. Remschmidt (1998). "Evidence for linkage of spelling disability to chromosome 15." Am J Hum Genet **63**(1): 279-282.

Schumacher, J., H. Anthoni, F. Dahdouh, I. R. Konig, A. M. Hillmer, N. Kluck, M. Manthey, E. Plume, A. Warnke, H. Remschmidt*, et al.* (2006). "Strong genetic evidence of DCDC2 as a susceptibility gene for dyslexia." Am J Hum Genet **78**(1): 52-62.

Schumacher, J., I. R. Konig, T. Schroder, M. Duell, E. Plume, P. Propping, A. Warnke, C. Libertus, A. Ziegler, B. Muller-Myhsok*, et al.* (2008). "Further evidence for a susceptibility locus contributing to reading disability on chromosome 15q15-q21." Psychiatr Genet **18**(3): 137-142.

Scott-Van Zeeland, A. A., B. S. Abrahams, A. I. Alvarez-Retuerto, L. I. Sonnenblick, J. D. Rudie, D. Ghahremani, J. A. Mumford, R. A. Poldrack, M. Dapretto, D. H. Geschwind*, et al.* (2010). "Altered functional connectivity in frontal lobe circuits is associated with variation in the autism risk gene CNTNAP2." Sci Transl Med **2**(56): 56ra80.

Sebat, J., B. Lakshmi, D. Malhotra, J. Troge, C. Lese-Martin, T. Walsh, B. Yamrom, S. Yoon, A. Krasnitz, J. Kendall*, et al.* (2007). "Strong association of de novo copy number mutations with autism." Science **316**(5823): 445-449.

Semenza, C., S. Bonollo, R. Polli, C. Busana, R. Pignatti, T. Iuculano, A. Maria Laverda, K. Priftis and A. Murgia (2012). "Genetics and mathematics: FMR1 premutation female carriers." Neuropsychologia **50**(14): 3757-3763.
Seshadri, S., A. L. DeStefano, R. Au, J. M. Massaro, A. S. Beiser, M. Kelly-Hayes, C. S. Kase, R. B. D'Agostino, Sr., C. Decarli, L. D. Atwood*, et al.* (2007). "Genetic correlates of brain aging on MRI and cognitive test measures: a genome-wide association and linkage analysis in the Framingham Study." BMC Med Genet **8 Suppl 1**: S15.
Sharma, S., X. Gao, D. Londono, S. E. Devroy, K. N. Mauldin, J. T. Frankel, J. M. Brandon, D. Zhang, Q. Z. Li, M. B. Dobbs*, et al.* (2011). "Genome-wide association studies of adolescent idiopathic scoliosis suggest candidate susceptibility genes." Hum Mol Genet **20**(7): 1456-1466.
Sharp, A. J., H. C. Mefford, K. Li, C. Baker, C. Skinner, R. E. Stevenson, R. J. Schroer, F. Novara, M. De Gregori, R. Ciccone*, et al.* (2008). "A recurrent 15q13.3 microdeletion syndrome associated with mental retardation and seizures." Nat Genet **40**(3): 322-328.
Shaywitz, S. E. and B. A. Shaywitz (2005). "Dyslexia (specific reading disability)." Biol Psychiatry **57**(11): 1301-1309.
Sherman, G. F., A. M. Galaburda and N. Geschwind (1985). "Cortical anomalies in brains of New Zealand mice: a neuropathologic model of dyslexia?" Proc Natl Acad Sci U S A **82**(23): 8072-8074.
Shinawi, M., C. P. Schaaf, S. S. Bhatt, Z. Xia, A. Patel, S. W. Cheung, B. Lanpher, S. Nagl, H. S. Herding, C. Nevinny-Stickel*, et al.* (2009). "A small recurrent deletion within 15q13.3 is associated with a range of neurodevelopmental phenotypes." Nat Genet **41**(12): 1269-1271.
Silverman, J. L., M. Yang, C. Lord and J. N. Crawley (2010). "Behavioural phenotyping assays for mouse models of autism." Nat Rev Neurosci **11**(7): 490-502.
Simonoff, E., A. Pickles, T. Charman, S. Chandler, T. Loucas and G. Baird (2008). "Psychiatric disorders in children with autism spectrum disorders: prevalence, comorbidity, and associated factors in a population-derived sample." J Am Acad Child Adolesc Psychiatry **47**(8): 921-929.
Simpson, N. H., L. Addis, W. M. Brandler, V. Slonims, A. Clark, J. Watson, T. S. Scerri, E. R. Hennessy, P. F. Bolton, G. Conti-Ramsden*, et al.* (2013). "Increased prevalence of sex chromosome aneuploidies in specific language impairment and dyslexia." Dev Med Child Neurol.
Sivertsen, B., M. B. Posserud, C. Gillberg, A. J. Lundervold and M. Hysing (2012). "Sleep problems in children with autism spectrum problems: a longitudinal population-based study." Autism **16**(2): 139-150.
Skafidas, E., R. Testa, D. Zantomio, G. Chana, I. P. Everall and C. Pantelis (2012). "Predicting the diagnosis of autism spectrum disorder using gene pathway analysis." Mol Psychiatry.
Smith, S. D., W. J. Kimberling, B. F. Pennington and H. A. Lubs (1983). "Specific reading disability: identification of an inherited form through linkage analysis." Science **219**(4590): 1345-1347.
Somerville, M. J., C. B. Mervis, E. J. Young, E. J. Seo, M. del Campo, S. Bamforth, E. Peregrine, W. Loo, M. Lilley, L. A. Perez-Jurado*, et al.* (2005). "Severe expressive-language delay related to duplication of the Williams-Beuren locus." N Engl J Med **353**(16): 1694-1701.
State, M. W. (2011). "The erosion of phenotypic specificity in psychiatric genetics: emerging lessons from CNTNAP2." Biol Psychiatry **69**(9): 816-817.

Stefansson, H., D. Rujescu, S. Cichon, O. P. Pietilainen, A. Ingason, S. Steinberg, R. Fossdal, E. Sigurdsson, T. Sigmundsson, J. E. Buizer-Voskamp*, et al.* (2008). "Large recurrent microdeletions associated with schizophrenia." Nature **455**(7210): 232-236.

Stein, M. B., B. Z. Yang, D. A. Chavira, C. A. Hitchcock, S. C. Sung, E. Shipon-Blum and J. Gelernter (2011). "A common genetic variant in the neurexin superfamily member CNTNAP2 is associated with increased risk for selective mutism and social anxiety-related traits." Biol Psychiatry **69**(9): 825-831.

Stogmann, E., E. Reinthaler, S. Eltawil, M. A. El Etribi, M. Hemeda, N. El Nahhas, A. M. Gaber, A. Fouad, S. Edris, A. Benet-Pages*, et al.* (2013). "Autosomal recessive cortical myoclonic tremor and epilepsy: association with a mutation in the potassium channel associated gene CNTN2." Brain **136**(Pt 4): 1155-1160.

Strauss, K. A., E. G. Puffenberger, M. J. Huentelman, S. Gottlieb, S. E. Dobrin, J. M. Parod, D. A. Stephan and D. H. Morton (2006). "Recessive symptomatic focal epilepsy and mutant contactin-associated protein-like 2." N Engl J Med **354**(13): 1370-1377.

Sutcliffe, J. S., Y. H. Jiang, R. J. Galijaard, T. Matsuura, P. Fang, T. Kubota, S. L. Christian, J. Bressler, B. Cattanach, D. H. Ledbetter*, et al.* (1997). "The E6-Ap ubiquitin-protein ligase (UBE3A) gene is localized within a narrowed Angelman syndrome critical region." Genome Res **7**(4): 368-377.

Szafranski, P., C. P. Schaaf, R. E. Person, I. B. Gibson, Z. Xia, S. Mahadevan, J. Wiszniewska, C. A. Bacino, S. Lalani, L. Potocki*, et al.* (2010). "Structures and molecular mechanisms for common 15q13.3 microduplications involving CHRNA7: benign or pathological?" Hum Mutat **31**(7): 840-850.

Szatmari, P., A. D. Paterson, L. Zwaigenbaum, W. Roberts, J. Brian, X. Q. Liu, J. B. Vincent, J. L. Skaug, A. P. Thompson, L. Senman*, et al.* (2007). "Mapping autism risk loci using genetic linkage and chromosomal rearrangements." Nat Genet **39**(3): 319-328.

Tabuchi, K., J. Blundell, M. R. Etherton, R. E. Hammer, X. Liu, C. M. Powell and T. C. Südhof (2007). "A neuroligin-3 mutation implicated in autism increases inhibitory synaptic transmission in mice." Science **318**(5847): 71-76.

Taipale, M., N. Kaminen, J. Nopola-Hemmi, T. Haltia, B. Myllyluoma, H. Lyytinen, K. Muller, M. Kaaranen, P. J. Lindsberg, K. Hannula-Jouppi*, et al.* (2003). "A candidate gene for developmental dyslexia encodes a nuclear tetratricopeptide repeat domain protein dynamically regulated in brain." Proc Natl Acad Sci U S A **100**(20): 11553-11558.

Tan, G. C., T. F. Doke, J. Ashburner, N. W. Wood and R. S. Frackowiak (2010). "Normal variation in fronto-occipital circuitry and cerebellar structure with an autism-associated polymorphism of CNTNAP2." Neuroimage **53**(3): 1030-1042.

Tapia-Paez, I., K. Tammimies, S. Massinen, A. L. Roy and J. Kere (2008). "The complex of TFII-I, PARP1, and SFPQ proteins regulates the DYX1C1 gene implicated in neuronal migration and dyslexia." FASEB J **22**(8): 3001-3009.

Terracciano, A., S. Sanna, M. Uda, B. Deiana, G. Usala, F. Busonero, A. Maschio, M. Scally, N. Patriciu, W. M. Chen*, et al.* (2010). "Genome-wide association scan for five major dimensions of personality." Mol Psychiatry **15**(6): 647-656.

The Dutch-Belgian Fragile X Consortium (1994). "Fmr1 knockout mice: a model to study fragile X mental retardation." Cell **78**(1): 23-33.

Toma, C., A. Hervas, B. Torrico, N. Balmana, M. Salgado, M. Maristany, E. Vilella, R. Martinez-Leal, M. I. Planelles, I. Cusco*, et al.* (2013). "Analysis of two language-related genes in autism: a case-control association study of FOXP2 and CNTNAP2." Psychiatr Genet **23**(2): 82-85.

Toro, R., M. Konyukh, R. Delorme, C. Leblond, P. Chaste, F. Fauchereau, M. Coleman, M. Leboyer, C. Gillberg and T. Bourgeron (2010). "Key role for gene dosage and synaptic homeostasis in autism spectrum disorders." Trends Genet **26**(8): 363-372.

Traka, M., L. Goutebroze, N. Denisenko, M. Bessa, A. Nifli, S. Havaki, Y. Iwakura, F. Fukamauchi, K. Watanabe, B. Soliven*, et al.* (2003). "Association of TAG-1 with Caspr2 is essential for the molecular organization of juxtaparanodal regions of myelinated fibers." J Cell Biol **162**(6): 1161-1172.

Turic, D., L. Robinson, M. Duke, D. W. Morris, V. Webb, M. Hamshere, C. Milham, E. Hopkin, K. Pound, S. Fernando*, et al.* (2003). "Linkage disequilibrium mapping provides further evidence of a gene for reading disability on chromosome 6p21.3-22." Mol Psychiatry **8**(2): 176-185.

Tzenova, J., B. J. Kaplan, T. L. Petryshen and L. L. Field (2004). "Confirmation of a dyslexia susceptibility locus on chromosome 1p34-p36 in a set of 100 Canadian families." Am J Med Genet B Neuropsychiatr Genet **127B**(1): 117-124.

Ucker, D. S. and K. R. Yamamoto (1984). "Early events in the stimulation of mammary tumor virus RNA synthesis by glucocorticoids. Novel assays of transcription rates." J Biol Chem **259**(12): 7416-7420.

Ueda, S., S. Fujimoto, K. Hiramoto, M. Negishi and H. Katoh (2008). "Dock4 regulates dendritic development in hippocampal neurons." J Neurosci Res **86**(14): 3052-3061.

van Abel, D., O. Michel, R. Veerhuis, M. Jacobs, M. van Dijk and C. B. Oudejans (2012). "Direct downregulation of CNTNAP2 by STOX1A is associated with Alzheimer's disease." J Alzheimers Dis **31**(4): 793-800.

van Bon, B. W., H. C. Mefford, B. Menten, D. A. Koolen, A. J. Sharp, W. M. Nillesen, J. W. Innis, T. J. de Ravel, C. L. Mercer, M. Fichera*, et al.* (2009). "Further delineation of the 15q13 microdeletion and duplication syndromes: a clinical spectrum varying from non-pathogenic to a severe outcome." J Med Genet **46**(8): 511-523.

Van der Aa, N., L. Rooms, G. Vandeweyer, J. van den Ende, E. Reyniers, M. Fichera, C. Romano, B. Delle Chiaie, G. Mortier, B. Menten*, et al.* (2009). "Fourteen new cases contribute to the characterization of the 7q11.23 microduplication syndrome." Eur J Med Genet **52**(2-3): 94-100.

van der Zwaag, B., W. G. Staal, R. Hochstenbach, M. Poot, H. A. Spierenburg, M. V. de Jonge, N. E. Verbeek, R. van 't Slot, M. A. van Es, F. J. Staal*, et al.* (2010). "A co-segregating microduplication of chromosome 15q11.2 pinpoints two risk genes for autism spectrum disorder." Am J Med Genet B Neuropsychiatr Genet **153B**(4): 960-966.

Veerappa, A. M., M. Saldanha, P. Padakannaya and N. B. Ramachandra (2013). "Family-based genome-wide copy number scan identifies five new genes of dyslexia involved in dendritic spinal plasticity." J Hum Genet **58**(8): 539-547.

Velayos-Baeza, A., C. Toma, S. da Roza, S. Paracchini and A. P. Monaco (2007). "Alternative splicing in the dyslexia-associated gene KIAA0319." Mamm Genome **18**(9): 627-634.

Velayos-Baeza, A., C. Toma, S. Paracchini and A. P. Monaco (2008). "The dyslexia-associated gene KIAA0319 encodes highly N- and O-glycosylated plasma membrane and secreted isoforms." Hum Mol Genet **17**(6): 859-871.

Veltman, M. W., E. E. Craig and P. F. Bolton (2005). "Autism spectrum disorders in Prader-Willi and Angelman syndromes: a systematic review." Psychiatr Genet **15**(4): 243-254.

Verhoeven, W. M., J. I. Egger, M. H. Willemsen, G. J. de Leijer and T. Kleefstra (2012). "Phelan-McDermid syndrome in two adult brothers: atypical bipolar disorder as its psychopathological phenotype?" Neuropsychiatr Dis Treat **8**: 175-179.

Verkerk, A. J., C. A. Mathews, M. Joosse, B. H. Eussen, P. Heutink and B. A. Oostra (2003). "CNTNAP2 is disrupted in a family with Gilles de la Tourette syndrome and obsessive compulsive disorder." Genomics **82**(1): 1-9.

Vernes, S. C., D. F. Newbury, B. S. Abrahams, L. Winchester, J. Nicod, M. Groszer, M. Alarcon, P. L. Oliver, K. E. Davies, D. H. Geschwind*, et al.* (2008). "A functional genetic link between distinct developmental language disorders." N Engl J Med **359**(22): 2337-2345.

Vorstman, J. A., W. G. Staal, E. van Daalen, H. van Engeland, P. F. Hochstenbach and L. Franke (2006). "Identification of novel autism candidate regions through analysis of reported cytogenetic abnormalities associated with autism." Mol Psychiatry **11**(1): 1, 18-28.

Vucurovic, K., E. Landais, C. Delahaigue, J. Eutrope, A. Schneider, C. Leroy, H. Kabbaj, J. Motte, D. Gaillard, A. C. Rolland*, et al.* (2012). "Bipolar affective disorder and early dementia onset in a male patient with SHANK3 deletion." Eur J Med Genet **55**(11): 625-629.

Waddington, C. H. (1942). "Canalization of development and the inheritance of acquired characters." Nature **150**: 563-565.

Wang, K., H. Zhang, D. Ma, M. Bucan, J. T. Glessner, B. S. Abrahams, D. Salyakina, M. Imielinski, J. P. Bradfield, P. M. Sleiman*, et al.* (2009). "Common genetic variants on 5p14.1 associate with autism spectrum disorders." Nature **459**(7246): 528-533.

Wang, K. S., X. F. Liu and N. Aragam (2010). "A genome-wide meta-analysis identifies novel loci associated with schizophrenia and bipolar disorder." Schizophr Res **124**(1-3): 192-199.

Weiss, L. A., D. E. Arking, M. J. Daly and A. Chakravarti (2009). "A genome-wide linkage and association scan reveals novel loci for autism." Nature **461**(7265): 802-808.

Weiss, L. A., Y. Shen, J. M. Korn, D. E. Arking, D. T. Miller, R. Fossdal, E. Saemundsen, H. Stefansson, M. A. Ferreira, T. Green*, et al.* (2008). "Association between microdeletion and microduplication at 16p11.2 and autism." N Engl J Med **358**(7): 667-675.

Whalley, H. C., G. O'Connell, J. E. Sussmann, A. Peel, A. C. Stanfield, M. E. Hayiou-Thomas, E. C. Johnstone, S. M. Lawrie, A. M. McIntosh and J. Hall (2011). "Genetic variation in CNTNAP2 alters brain function during linguistic processing in healthy individuals." Am J Med Genet B Neuropsychiatr Genet **156B**(8): 941-948.

Wheelwright, S., B. Auyeung, C. Allison and S. Baron-Cohen (2010). "Defining the broader, medium and narrow autism phenotype among parents using the Autism Spectrum Quotient (AQ)." Mol Autism **1**(1): 10.

Whitehouse, A. J., D. V. Bishop, Q. W. Ang, C. E. Pennell and S. E. Fisher (2011). "CNTNAP2 variants affect early language development in the general population." Genes Brain Behav **10**(4): 451-456.

Whitehouse, A. J., H. J. Watt, E. A. Line and D. V. Bishop (2009). "Adult psychosocial outcomes of children with specific language impairment, pragmatic language impairment and autism." Int J Lang Commun Disord **44**(4): 511-528.

Wigg, K. G., J. M. Couto, Y. Feng, B. Anderson, T. D. Cate-Carter, F. Macciardi, R. Tannock, M. W. Lovett, T. W. Humphries and C. L. Barr (2004). "Support for EKN1 as the susceptibility locus for dyslexia on 15q21." Mol Psychiatry **9**(12): 1111-1121.

Wilcke, A., C. Ligges, J. Burkhardt, M. Alexander, C. Wolf, E. Quente, P. Ahnert, P. Hoffmann, A. Becker, B. Muller-Myhsok*, et al.* (2012). "Imaging genetics of FOXP2 in dyslexia." Eur J Hum Genet **20**(2): 224-229.
Wilcke, A., J. Weissfuss, H. Kirsten, G. Wolfram, J. Boltze and P. Ahnert (2009). "The role of gene DCDC2 in German dyslexics." Ann Dyslexia **59**(1): 1-11.
Willcutt, E. G. and B. F. Pennington (2000). "Psychiatric comorbidity in children and adolescents with reading disability." J Child Psychol Psychiatry **41**(8): 1039-1048.
Wing, L. (1981). "Asperger's syndrome: a clinical account." Psychol Med **11**(1): 115-129.
Wolff, P. H., J. Gardner, J. Lappen, J. Paccia and D. Meryash (1988). "Variable expression of the fragile X syndrome in heterozygous females of normal intelligence." Am J Med Genet **30**(1-2): 213-225.
Woltereck, R. (1909). "Weitere experimentelle Untersuchungen über Artveränderung, speziel über das Wesen quantitativer Artunterschiede bei Daphniden." Verhandlungen der deutschen zoologischen Gesellschaft **19** 110–173.
Won, H., H.-R. Lee, H. Y. Gee, W. Mah, J.-I. Kim, J. Lee, S. Ha, C. Chung, E. S. Jung, Y. S. Cho*, et al.* (2012). "Autistic-like social behaviour in Shank2-mutant mice improved by restoring NMDA receptor function." Nature **486**(7402): 261-265.
Worthey, E. A., G. Raca, J. J. Laffin, B. M. Wilk, J. M. Harris, K. J. Jakielski, D. P. Dimmock, E. A. Strand and L. D. Shriberg (2013). "Whole-exome sequencing supports genetic heterogeneity in childhood apraxia of speech." J Neurodev Disord **5**(1): 29.
Y., L. I. (1973). "S egmentation of the spoken word and reading acquisition." Bulletin of the Orton Society **23**: 65-77.
Yamasaki, K., K. Joh, T. Ohta, H. Masuzaki, T. Ishimaru, T. Mukai, N. Niikawa, M. Ogawa, J. Wagstaff and T. Kishino (2003). "Neurons but not glial cells show reciprocal imprinting of sense and antisense transcripts of Ube3a." Hum Mol Genet **12**(8): 837-847.
Yang, J., B. Benyamin, B. P. McEvoy, S. Gordon, A. K. Henders, D. R. Nyholt, P. A. Madden, A. C. Heath, N. G. Martin, G. W. Montgomery*, et al.* (2010). "Common SNPs explain a large proportion of the heritability for human height." Nat Genet **42**(7): 565-569.
Yaspan, B. L., W. S. Bush, E. S. Torstenson, D. Ma, M. A. Pericak-Vance, M. D. Ritchie, J. S. Sutcliffe and J. L. Haines (2011). "Genetic analysis of biological pathway data through genomic randomization." Hum Genet **129**(5): 563-571.
Yu, T. W., M. H. Chahrour, M. E. Coulter, S. Jiralerspong, K. Okamura-Ikeda, B. Ataman, K. Schmitz-Abe, D. A. Harmin, M. Adli, A. N. Malik*, et al.* (2013). "Using whole-exome sequencing to identify inherited causes of autism." Neuron **77**(2): 259-273.
Zilina, O., T. Reimand, P. Zjablovskaja, K. Mannik, M. Mannamaa, A. Traat, H. Puusepp-Benazzouz, A. Kurg and K. Ounap (2012). "Maternally and paternally inherited deletion of 7q31 involving the FOXP2 gene in two families." Am J Med Genet A **158A**(1): 254-256.
Zweier, C., E. K. de Jong, M. Zweier, A. Orrico, L. B. Ousager, A. L. Collins, E. K. Bijlsma, M. A. Oortveld, A. B. Ekici, A. Reis*, et al.* (2009). "CNTNAP2 and NRXN1 are mutated in autosomal-recessive Pitt-Hopkins-like mental retardation and determine the level of a common synaptic protein in Drosophila." Am J Hum Genet **85**(5): 655-666.

Annexes

The Emerging Role of *SHANK* Genes in Neuropsychiatric Disorders

Audrey Guilmatre,[1,2,3] Guillaume Huguet,[1,2,3] Richard Delorme,[1,2,3,4] Thomas Bourgeron[1,2,3]

[1] Human Genetics and Cognitive Functions Unit, Institut Pasteur, Paris, France

[2] CNRS URA 2182 'Genes, Synapses and Cognition,' Institut Pasteur, Paris, France

[3] Human Genetics and Cognitive Functions, University Paris Diderot, Sorbonne Paris Cité, Paris, France

[4] Child and Adolescent Psychiatry, Robert Debré Hospital, Paris, Paris

Received 28 March 2013; accepted 28 August 2013

ABSTRACT: The genetic heterogeneity of neuropsychiatric disorders is high, but some pathways emerged, notably synaptic functioning. A large number of mutations have been described in genes such as neuroligins, neurexins, and *SHANK* that play a role in the formation and the maintenance of synapses. This review focuses on the disorders associated with mutations in *SHANK3* and the other members of its family, *SHANK1* and *SHANK2*. SHANKs are scaffolding proteins of the postsynaptic density of glutamatergic synapses. *SHANK3* has been described in the Phelan-McDermid syndrome (PMS), but also in autism spectrum disorders (ASD) and schizophrenia associated to moderate to severe intellectual disability (ID) and poor language. The evolution of patients with PMS includes symptoms of bipolar disorder and regression. *SHANK2* has been identified in patients with ASD with mild to severe ID. *SHANK1* has been associated with high-functioning autism in male patients, while carrier females only display anxiety and shyness. Finally, based on neuropathological findings in animal models and patients, a possible role of SHANK in Alzheimer's disease is discussed. Altogether, this review describes the clinical trajectories associated with different mutations of the *SHANK* genes and provides information to further investigate the role of the *SHANK* genes in neuropsychiatric disorders.
Develop Neurobiol 00: 000–000, 2013

Keywords: **SHANK; neuropsychiatric disorders; Phelan-McDermid syndrome**

INTRODUCTION

The genetic landscapes of neuropsychiatric disorders are complex and highly heterogeneous. Due to the absence of classical Mendelian segregation, the initial genetic models for these complex traits involved many genes with low effect. More recently, however, the identification of apparently monogenic forms of autism spectrum disorders (ASD) or schizophrenia was in favor of a combination of low effect alleles and rare mutations with medium to high penetrance. During the last decade, many neuropsychiatric disorders have been associated with genetic mutations or genomic imbalance known as copy number variations. To date, the causative genes usually account for less than 1% of the patients, but remarkably the same biological pathways are emerging in different neuropsychiatric disorders such as ASD and

Correspondence to: T. Bourgeron (thomasb@pasteur.fr).
Contract grant sponsors: ANR; contract grant number: ANR-08-MNPS-037-01—SynGen.

Contract grant sponsors: Institut Pasteur, the Bettencourt-Schueller Foundation, CNRS, University Paris Diderot, Neuron-ERANET (EUHF-AUTISM), the Fondation de France, the Conny-Maeva Charitable Foundation, the Cognacq Jay Foundation, the Orange Foundation, and the Fondamental Foundation.

Published online 00 Month 2013 in Wiley Online Library (wileyonlinelibrary.com).
DOI 10.1002/dneu.22128

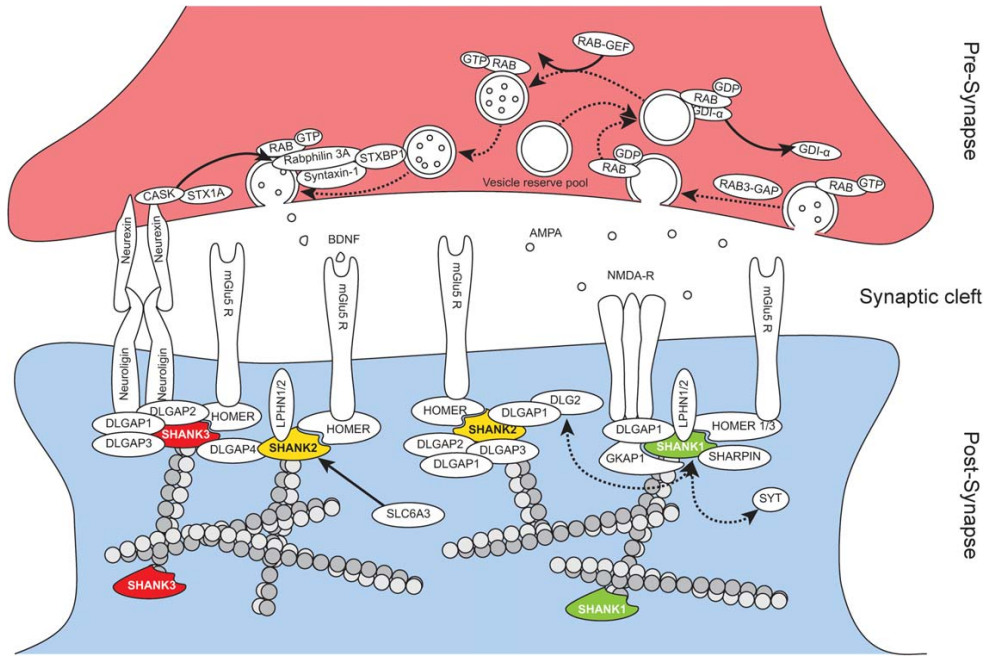


Figure 1 Interaction network of the SHANK at the post-synaptic density of glutamatergic synapses.

schizophrenia. Among these biological pathways, one of the most studied is composed of synaptic genes such as neuroligins, neurexins, and *SHANK* (Jamain et al., 2003; Durand et al., 2007; Toro et al., 2010). These proteins are crucial for synapse formation and maintenance as well as for correct excitatory and inhibitory balance (Sudhof, 2008). In this review, we will focus on the role of the *SHANK* gene family in neuropsychiatric disorders.

There are three *SHANK* genes in the human genome that are expressed in different regions of the brain. *SHANK* genes code for large synaptic scaffold proteins and bind to a variety of other proteins at the post-synaptic density (PSD) of excitatory synapses (Naisbitt et al., 1999) (Fig. 1). The *SHANK1* gene is located on chromosome 19q13.33, spans 55.1 kb and contains 23 exons and two alternative promoters leading to two distinct isoforms. It is highly expressed in mouse cerebral cortex and hippocampus (Lim et al., 1999; Bockers et al., 2004). The *SHANK2* gene is located on chromosome 11q13.3, and spans 621.8 kb. It contains 25 exons, three alternative promoters and one alternative stop codon leading to four distinct isoforms (Du et al., 1998; Boeckers et al., 1999). In mouse, *Shank2* is broadly expressed in the brain with a high expression in cerebellar Purkinje cells (Boeckers et al., 1999; Bockers et al., 2004; Peca et al., 2011; Schmeisser et al., 2012). The *SHANK3* gene is located on chromosome 22q13.3 and spans 55.1 kb. It contains 24 exons and one alternative stop codon located in the exon 21b leading to three distinct isoforms. It is highly expressed in mouse thalamus and striatum and in cerebellar granule cells (Boeckers et al., 1999; Naisbitt et al., 1999; Durand et al., 2007; Peca et al., 2011; Ey et al., 2012).

SHANK proteins assemble into large molecular platforms at the PSD interconnecting proteins of the postsynaptic membrane with the actin cytoskeleton of the dendritic spine. SHANK2 is among the first proteins to be expressed at the developing PSD followed by SHANK3. Major interaction partners include members of the NMDA receptor complex, members of the metabotropic glutamate receptor complex and actin-associated proteins (Fig. 1) (Boeckers et al., 2002; Grabrucker et al., 2011; Verpelli et al., 2012). The role of SHANK proteins has been investigated in rodents. *Shank1* knock-out mice display increased anxiety, decreased vocal communication, decreased locomotion and remarkably, enhanced working memory, but decreased long-term memory (Hung et al., 2008; Silverman et al., 2011; Wohr et al., 2011). *Shank2* knock-out mice present with hyperactivity, increased anxiety, repetitive grooming, and abnormalities in vocal and social behaviors (Schmeisser et al., 2012; Won et al., 2012). *Shank3* knock-out mice show self-injurious repetitive

grooming, and deficits in social interaction and communication (Bozdagi et al., 2010; Peca et al., 2011; Wang et al., 2011; Yang et al., 2012).

In humans, the *SHANK3* gene was the first to be studied notably because of its role in Phelan-McDermid syndrome (PMS) (see below), but more recently, *SHANK1* and *SHANK2* have also been implicated in diverse neuropsychiatric conditions. This review summarizes the genetic and clinical findings of studies investigating the role of *SHANK* genes in a broad range of clinical conditions including PMS, ASD, schizophrenia, bipolar disorders, and Alzheimer disease (Fig. 2).

PHELAN-McDERMID SYNDROME

The PMS is a contiguous gene syndrome, caused by a 22q13 deletion encompassing several genes such as *ACR*, *RABL2B*, *MGC70863*, and *SHANK3* (for review see (Phelan and McDermid, 2012)). The first case of 22q13 deletion syndrome has been reported in 1985 in a child with severe intellectual disability (ID), absence of speech and facial dysmorphism, in a family with a pericentric inversion at chromosome 22 leading to a mosaic partial trisomy 22 in his mother and siblings (Watt et al., 1985). Nesslinger et al. (1994) reported additional cases and suggested that it might be a recognizable syndrome. Phelan et al. (2001) reported a comprehensive description of 37 cases that led to the definition of the PMS. To date, over 600 cases have been reported leading to the detailed clinical description of the syndrome. More than 75% of cases with PMS display global developmental delays, absent or severely delayed speech, normal or accelerated growth, neonatal hypotonia, and facial dysmorphism (Phelan and McDermid, 2012). ASD have been described in approximately half of the patients. Seizures were also described in >25% of cases, such as grand mal, focal and absence seizures, and continuous spike and waves during slow-wave sleep syndrome. To date, no specific electroencephalogram profile was associated with PMS (Lesca et al., 2012; Phelan and McDermid, 2012).

Deletion sizes are variable ranging from <100 kb to >9 Mb. Except for two patients (Wilson et al., 2008), all patients with PMS have a 22q13 deletion leading to the haploinsufficiency of four genes: *SHANK3*, *ACR*, *RABL2B*, and *MGC70863*. Small deletions of 140 kb were described in several cases with a similar breakpoint between exons 7 and 8 of *SHANK3* (Bonaglia et al., 2006; Durand et al., 2007). The breakpoint is located in a short genomic region, highly rich in guanine and with a high probability of being in a G-quartet conformation (Durand et al., 2007). The G-quartet structure seems to be more prone to genomic rearrangements than the classical structure of double helix DNA (Durand et al., 2007). The patients carrying these small deletions present with mild ID, speech delay, autistic features, and psychomotor epilepsy (Flint et al., 1995; Anderlid et al., 2002).

The first evidence of *SHANK3* as a main gene associated with the neuropsychiatric features of patients with PMS was the description of a patient with a *de novo* reciprocal balanced translocation t(12;22)(q24.1;q13.3) interrupting *SHANK3* in exon 21 (Bonaglia et al., 2001). His phenotype was typical for patients with PMS: mild ID, severe speech delay, mild hypotonia, and minor facial dysmorphism. A (22;X) translocation encompassing the last 2 exons of *SHANK3* and part of *ACR* has been observed in a patient with a PMS like phenotype, further supporting the role of *SHANK3* in this syndrome (Misceo et al., 2011).

SHANK AND AUTISM SPECTRUM DISORDERS

Durand et al. (2007) described three patients with ASD, severe/profound ID, and language delay or absence of speech carrying *SHANK3* interstitial or terminal deletion or frameshift mutation. Interestingly, the brother of the patient carrying the 22qter deletion had the reciprocal 22qter duplication and was diagnosed with Asperger syndrome, suggesting a gene dosage effect of *SHANK3*. Moessner et al. subsequently screened 400 patients with ASD for variants in *SHANK3* and also found two families with a *de novo* deletion segregating with ASD, a point mutation, and a deletion in a proband associated to a reciprocal duplication in her sister with attention deficit and hyperactivity disorder (ADHD) (Moessner et al., 2007). These patients presented with ASD, moderate to severe ID, and poor or no language. Gauthier et al. (2009) reported a *de novo* splice site mutation in a patient with autism and a functionally deleterious L68P mutation in a patient with PDD-NOS, inherited from her father who suffered from epilepsy. Finally, Boccuto et al. (2013) further supported the involvement of *SHANK3* in ASD with the description of a group of patients carrying deleterious mutations (Boccuto et al., 2013). Such deleterious variants have also been reported in patients with ASD and severe delayed speech development in the Japanese population (Waga et al., 2011).

Beyond the recurrent involvement of *SHANK3* in ASD, several studies also identified deleterious mutations of *SHANK2* and *SHANK1*. Berkel et al. (2010) described three patients with ASD and ID with

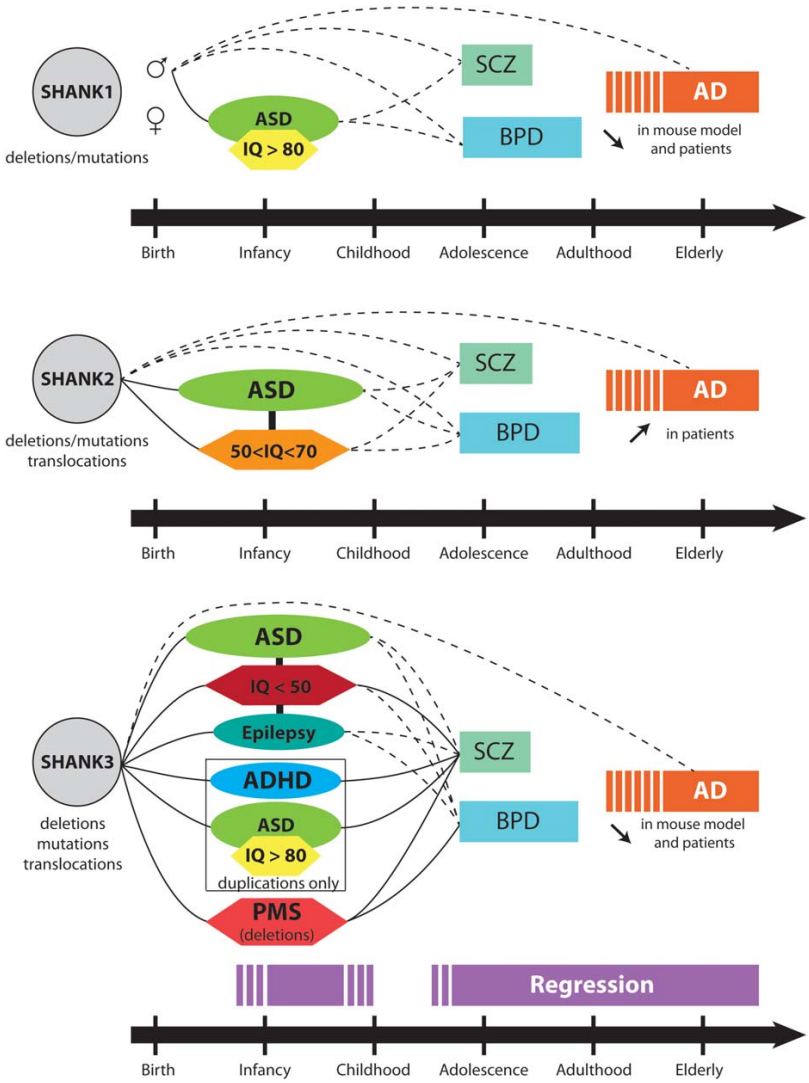


Figure 2 Clinical features and possible trajectories of patients carrying heterozygous deleterious mutations or copy number variations in the different *SHANK* genes. A: *SHANK1* deletions and mutations cause high-functioning autism in males. The T allele of SNP rs3810280 in the *SHANK1* promoter has been associated with reduced working memory in patients with SCZ. Decreased level of SHANK1 has been observed in brain of mouse models and patients with AD. B: *SHANK2* mutations/deletions/translocations are found in patients with ASD and mild to moderate intellectual disability. SHANK2 level is increased in brain of patients with AD. C: *SHANK3* mutations/deletions/translocations are associated with Phelan-McDermid syndrome, ASD, and SCZ with moderate to severe ID and epilepsy in >25% of cases. Bipolar disorder and regression was described in adult patients with Phelan-McDermid syndrome. Duplications were observed in patients with high-functioning autism, ADHD or SCZ. SHANK3 level is decreased in both brains of mouse models of AD and patients with AD. AD = Alzheimer's disease, ADHD = attention deficit hyperactivity disorder, ASD = autism spectrum disorders, BPD = bipolar disorders, IQ = intellectual quotient, SCZ = schizophrenia. Plain lines indicate phenotypic outcomes already associated with *SHANK* genes mutations and dotted lines show candidate disorders that should be screened for *SHANK* genes mutations.

de novo heterozygous *SHANK2* deletions. A group of seven patients with ASD and ID (one of them presenting with ID only) were also carriers of rare nonsynonymous mutations predicted as deleterious, but inherited from an unaffected parent. Five independent studies then reported additional *de novo* heterozygous *SHANK2* deletions in patients with ASD, speech, and developmental delay, and moderate ID (Pinto et al., 2010; Wischmeijer et al., 2011; Leblond et al., 2012; Schluth-Bolard et al., 2013; Chilian et al., 2013). Leblond et al. also observed an enrichment of mutations affecting conserved amino acids and altering synapse density in patients with ASD (Leblond et al., 2012). Interestingly, two patients with *SHANK2* deletion (Leblond et al., 2012) and a patient with a translocation disrupting *SHANK2* (Chilian et al., 2013) also carried a *CHRNA7* duplication and *ARHGAP11B* deletion at 15q11-q13. The low probability of such a co-occurrence suggested that an epistasis could exist between the *SHANK2* gene and genes on chromosome 15q11-q13. A third patient with a *SHANK2* deletion also carried a deletion at 15q11-q13 removing the *CYFIP1* gene that codes a partner of the fragile X protein FMRP (Leblond et al., 2012).

The role of *SHANK1* in ASD was first reported by Sato et al. (2012). An inherited deletion of 63.8 kb, encompassing *SHANK1* and the *CLEC11A* gene, was segregating in six individuals from the same family. Interestingly, the four males presented with high-functioning autism or broader autism phenotype, while the two females carrying the same deletion only displayed anxiety and shyness. These results suggested that females were protected from the impact of *SHANK1* deletions. In an independent family, a *de novo* 63.4 kb deletion encompassing *SYT3* and the last three exons of *SHANK1* was identified in a male patient with high-functioning autism. Based on these results, *SHANK1* deletions might be associated with high-functioning autism in males. As for *SHANK2*, multiple hits might be also present. A nonsense mutation in the protocadherin *PCDHGA11* (located on a different chromosome than *SHANK1*) was observed in the six family members carrying the *SHANK1* deletion and a deletion of the *SYT3* gene was also found in the sporadic case from the second family.

In summary, mutations of all *SHANK* genes were associated with ASD, but additional studies are warranted to better ascertain their prevalence and the clinical trajectories of the patients.

SHANK AND SCHIZOPHRENIA

In 2010, Gauthier et al. sequenced the coding regions of *SHANK3* in 185 patients with schizophrenia (Gauthier et al., 2010). They reported two *de novo* *SHANK3* mutations in two unrelated families. In one family, three brothers were carrying a R1117X *de novo* *SHANK3* truncating mutations most likely due to a germline mosaicism in the father. All three patients have been diagnosed with schizophrenia (schizoaffective disorder in the proband) and borderline to moderate ID. The older brother presented with ADHD in childhood and had one seizure episode at 10 years old. He was described as having a mild ID. The younger brother had atypical chronic psychosis and a moderate ID. None of them had dysmorphic features. In the second family, the proband was carrying a *de novo* R536W *SHANK3* mutation. She had a borderline ID (global intellectual quotient (IQ) of 73) in addition to a schizoaffective disorder. First schizoaffective symptoms emerged when he was 11. She had normal growth, no dysmorphic features, speech impairment, and poor academic and social performance.

In 2007, Failla et al. reported a patient with a 5.4 Mb 22q13.3-qter duplication encompassing *SHANK3* (Failla et al., 2007). She presented with a normal development until 13 years old. At this age, she started to display behavioral problems, with auditory hallucinations, loss of self-control and dysfunction of self-awareness, disorientation, followed by aggressiveness and sleep disturbances. She had a global IQ of 73 at age 20. She also displayed psychomotor restlessness, muscular hypotonia, and attention deficit, together with impaired ideation, incoherent speech, irritability, sexually related problems, unstable temper, and compromised general affectivity. She presented adaptive difficulties, apathy, and poor personal hygiene leading to the diagnosis of borderline intellectual functioning and disorganized schizophrenia. She presented with a facial dysmorphism, scoliosis, splayfoot with valgus hallux, and joint hyperlaxity.

To date, there is no report of *SHANK1* and *SHANK2* mutation screening in patients with schizophrenia. However, Lennertz et al. (2012) analyzed 5 *SHANK1* SNPs in 199 patients with schizophrenia and 206 controls and found a significant association between the T allele of SNP rs3810280 ($p < 0.001$, resisting Bonferroni correction) located in *SHANK1* promoter and reduced auditory working memory in patients with schizophrenia, which was replicated in 77 individuals at-risk for schizophrenia ($p = 0.044$). Interestingly, a previous pharmacological study showed effects of the antipsychotic drugs clozapine and haloperidol, used in the treatment of schizophrenia, on shank1 density (Critchlow et al., 2006). These results need replications on a larger cohort, but could point at one low effect allele affecting *SHANK1* expression.

SHANK AND BIPOLAR DISORDER

In 1996, Sovner et al. reported a 21-year-old patient with a ring chromosome 22 diagnosed with atypical bipolar disorder (Sovner et al., 1996). While there is no detailed molecular analysis, the patient phenotype was most likely a 22q13 deletion syndrome involving *SHANK3*. Following this initial publication, four independent studies reported the development of atypical bipolar disorder accompanied by a loss of skills in adult patients with PMS.

The two first studies reported the same family with two affected brothers with PMS (Verhoeven et al., 2012; Willemsen et al., 2012). Until the age of 17, the development of the younger brother was marked by the severity of the ID and hyperactivity together with temper tantrums and absence of language. At this age, in a couple of weeks, he began to lose weight, display sleep disturbances, but also changed his behavior with a loss of interest in daily activities, an increase of his social withdrawal and his repetitive behaviors with a marked anxiety. A severe major depressive disorder was first diagnosed followed by an atypical bipolar disorder. The elder brother also had a development delay but with a mild ID. At age 27, he presented with recurrent major depressive episodes and unstable pattern of mood. He was also diagnosed with atypical bipolar disorder, but less severe and with a later onset.

An additional case-report described four patients with PMS and bipolar disorder (Denayer et al., 2012). All individuals presented with severe to profound ID, ASD, self-absorbing behaviors and—for one of them—problems with social relationship, disruptive behaviors, and anxiety. The diagnosis of bipolar disorder was based on the presence of at least one manic episode with irritable mood, psychomotor agitation, decreased need for sleep, and increased babbling and talking, despite these patients otherwise did not talk. One patient was diagnosed with ADHD. One common feature of these patients is a loss of skills after adolescence or in early adulthood, such as deterioration in language, motor coordination, a progressive rigidity of the posture with shuffling gait, loss of the ability to do handwork and to eat with a knife and fork, a reduced eye contact, and diminished social interest. One patient was hospitalized after a probable benzodiazepine overdose leading to blood pressure fall and decreased consciousness followed by a catatonic period. After this episode, she lost her language, stopped interacting with others and did no longer recognize her mother, with increased mood swings. She stopped eating independently and also lost continence. Another bipolar patient developed a malignant neuroleptic syndrome during a substantial manic episode at 27 years old, after which he lost the ability to walk or eat independently, with a long recovery period. At 40 years, he underwent a septic shock and indefinitely lost language, walking, independent eating and dressing, and continence. One of the patients who suffered from epilepsy became totally dependent and bedridden after a prolonged epileptic state.

Vucurovic et al. also described a patient with a *de novo* complex multiple deletion encompassing at least six *SHANK3* exons and the genes *ACR*, *RABL2B*, and *RPL23AP82* (Vucurovic et al., 2012). At 3 years old, he displayed inattention, hyperactivity, and impulsive behavior with severe language delay. He was diagnosed with severe ID without autistic traits at 6 years old. He also developed a nonspecific sleep disorder. At 16 years old, he presented a depressive episode and social isolation, followed by psychomotor agitation and insomnia with impulsiveness and aggressiveness. He was hospitalized in inpatient psychiatric service where total insomnia with psychomotor agitation, anorexia, poor attention, disinhibited behavior, affected instability, euphoric, and then expensive mood with rapid depressive shifts were notably noticed. A diagnosis of affective bipolar disorder not otherwise specified was given. However, an early dementia was also hypothesized because of stereotyped behavior and regression of expressive speech and bladder control loss at 17 years old. In accordance with this, the protein profile observed in his cerebrospinal fluid was reminiscent to those of patients with Alzheimer's type dementia.

Finally, the first geriatric case with PMS was recently reported by Verhoeven et al. (2013). The early development of the patient was similar to the patients with PMS. During adulthood, she was diagnosed with atypical bipolar affective disorder. She had a mood dysregulation with disinhibited behaviors and sleep disturbances. She had an increasing tolerance to pain and her behavior gradually deteriorated in her 60's with severe sleep disturbances and gross motor perseveration.

SHANK AND ALZHEIMER'S DISEASE

A majority of patients with *SHANK3* deleterious mutations seems to experience a significant loss of skills during adulthood, but little is known on the clinical trajectories of elderly individuals carrying deleterious *SHANK* mutations. While there is currently no *SHANK* mutation screening reported in patients with Alzheimer's disease (AD), three

neurobiological studies pointed out a possible link between *SHANK* and AD pathophysiology. Amyloid-beta (Aβ) oligomers are markers of AD that accumulate in the forebrains of patients (Hardy and Selkoe, 2002). Following this observation, a mouse model for AD was made that overexpresses the amyloid precursor protein (APP) (Quon et al., 1991). Treatment of rat frontocortical neurons with the 1–40 Aβ isoform led to decreased synaptic levels of HOMER1B, mGluR1, and SHANK1 (Roselli et al., 2009). A similar decrease in SHANK1 and SHANK3 was observed in the brain of the APP transgenic mice (Pham et al., 2010). Finally, the same authors provided support for a decrease of SHANK1 level in the frontal cortex of patients with AD. Independently, Gong et al. (2009) observed an increase of SHANK2 levels in synaptosomes isolated from middle frontal gyrus in patients with AD as compared with controls. Remarkably, as mentioned previously, a low level of Aβ (479 mg/L) in the cerebrospinal fluid was also detected in a patient with bipolar disorders and carrying a *SHANK3* deletion.

Taken together, one can hypothesize that SHANK dysregulation might be part of the molecular pathology of AD and that patients with *SHANK* mutations could evolve with an AD. Future screening of *SHANK* genes in patients with AD should ascertain whether or not mutations in *SHANK* genes could be risk factors for AD.

DISCUSSION

Historically, the role of *SHANK* genes in neuropsychiatric disorders has been discovered with the description of patients with PMS carrying deletions that encompass *SHANK3*. Given the high frequency of autism in this disorder, *SHANK3* copy-number variants and mutations were assessed in large cohorts of patients with ASD. This strategy revealed to be successful. The emerging neuropsychiatric phenotype of patients with *SHANK3* deletions or mutations includes ASD, moderate to severe ID, and absence or delay of language, but also, less frequently, mood instability, delirium, hallucinations and, early or late cognitive regression. Following these results, *SHANK3* mutations were also described in cases with schizophrenia and ID and non-syndromic ID, suggesting that haploinsufficiency of *SHANK3* is a risk factor for a broad range of neuropsychiatric disorders.

Given the implication of *SHANK3* in these disorders and the existence of two other members of the family, attention was paid on *SHANK1* and *SHANK2*. *SHANK2* deletions and mutations have been described in ASD and mild to moderate ID, and non-syndromic ID, with a potential epistasis with *CHRNA7* duplication that co-occurred in independent cases of *SHANK2* deletions. Finally, *SHANK1* deletions have been reported in males with high-functioning autism, while the carrier females only display anxiety and shyness.

Altogether, a gradient of severity seems to appear depending on which *SHANK* gene is mutated. Apparently, mutations of *SHANK3* lead to a severe phenotype, while mutations in *SHANK2* or *SHANK1* are associated with a milder phenotype or even no psychiatric phenotype in females. The reason why the clinical phenotype varies from one *SHANK* gene to another remains largely unknown. The different patterns of expression in different brain regions and at different times during the development might partly explain the distinct phenotypic consequences. In addition, even within one group of patients with *SHANK3* mutations, the clinical trajectories of the patients can be very different (ID only, ASD, schizophrenia, and bipolar disorder) (Fig. 2). One of the common clinical features observed in patients with *SHANK3* mutations is the presence of ID. Why some patients will develop ASD or later in life schizophrenia or bipolar disorders remains unknown. The identification of additional genetic and/or environmental factors influencing the clinical trajectories will probably help to make a better diagnosis of these patients.

One key area of research that needs to be explored in more details in the future concerns the dysregulation of SHANK proteins in AD together with the very frequent regression undergone by adult patients carrying *SHANK3* mutations. Patients with mutations in a *SHANK* gene should therefore benefit from a follow-up when they reach adult age to monitor potential regression. Several genes such as *MAPT* on chromosome 17q21 are already known to play a role in both early onset psychoses and late dementias (Shaw-Smith et al., 2006). Deletions of *MAPT* cause ID, while duplication and mutations are involved in fronto-temporal dementia (Hutton et al., 1998; Rovelet-Lecrux et al., 2010).

In summary, since the first mutations identified in *SHANK* genes, there was a spectacular increase in our knowledge on the role of SHANK in neuropsychiatric disorders. We have a better idea on the clinical phenotypes associated with *SHANK* mutations. Mouse models for each Shank gene were developed to characterize the mechanisms involved in the synaptic and behavior deficits. In the very near future, human neuronal cells carrying *SHANK* mutations will be derived from induced pluripotent stem cells in order to

ascertain the synaptic deficits within the genomic background of the patients. This tight collaborative work between geneticists, neurobiologists, and clinicians should lead to a better diagnosis and to the identification of novel knowledge based treatments.

REFERENCES

Anderlid BM, Schoumans J, Anneren G, Tapia-Paez I, Dumanski J, Blennow E, Nordenskjold M. 2002. FISH-mapping of a 100-kb terminal 22q13 deletion. Hum Genet 110:439–443.

Berkel S, Marshall CR, Weiss B, Howe J, Roeth R, Moog U, Endris V, et al. 2010. Mutations in the SHANK2 synaptic scaffolding gene in autism spectrum disorder and mental retardation. Nat Genet 42:489–491.

Boccuto L, Lauri M, Sarasua SM, Skinner CD, Buccella D, Dwivedi A, Orteschi D, et al. 2013. Prevalence of SHANK3 variants in patients with different subtypes of autism spectrum disorders. Eur J Hum Genet 21:310–316.

Bockers TM, Segger-Junius M, Iglauer P, Bockmann J, Gundelfinger ED, Kreutz MR, Richter D, et al. 2004. Differential expression and dendritic transcript localization of Shank family members: Identification of a dendritic targeting element in the 3′ untranslated region of Shank1 mRNA. Mol Cell Neurosci 26:182–190.

Boeckers TM, Bockmann J, Kreutz MR, Gundelfinger ED. 2002. ProSAP/Shank proteins—A family of higher order organizing molecules of the postsynaptic density with an emerging role in human neurological disease. J Neurochem 81:903–910.

Boeckers TM, Winter C, Smalla KH, Kreutz MR, Bockmann J, Seidenbecher C, Garner CC, et al. 1999. Proline-rich synapse-associated proteins ProSAP1 and ProSAP2 interact with synaptic proteins of the SAPAP/GKAP family. Biochem Biophys Res Commun 264: 247–252.

Bonaglia MC, Giorda R, Borgatti R, Felisari G, Gagliardi C, Selicorni A, Zuffardi O. 2001. Disruption of the ProSAP2 gene in a t(12;22)(q24.1;q13.3) is associated with the 22q13.3 deletion syndrome. Am J Hum Genet 69:261–268.

Bonaglia MC, Giorda R, Mani E, Aceti G, Anderlid BM, Baroncini A, Pramparo T, et al. 2006. Identification of a recurrent breakpoint within the SHANK3 gene in the 22q13.3 deletion syndrome. J Med Genet 43:822–828.

Bozdagi O, Sakurai T, Papapetrou D, Wang X, Dickstein DL, Takahashi N, Kajiwara Y, et al. 2010. Haploinsufficiency of the autism-associated Shank3 gene leads to deficits in synaptic function, social interaction, and social communication. Mol Autism 1:1–5.

Chilian B, Abdollahpour H, Bierhals T, Haltrich I, Fekete G, Nagel I, Rosenberger G, et al. 2013. Dysfunction of SHANK2 and CHRNA7 in a patient with intellectual disability and language impairment supports genetic epistasis of the two loci. Clin Genet. [Epub ahead of print]

Critchlow HM, Maycox PR, Skepper JN, Krylova O. 2006. Clozapine and haloperidol differentially regulate dendritic spine formation and synaptogenesis in rat hippocampal neurons. Mol Cell Neurosci 32:356–365.

Denayer A, Van Esch H, de Ravel T, Frijns JP, Van Buggenhout G, Vogels A, Devriendt K, et al. 2012. Neuropsychopathology in 7 patients with the 22q13 deletion syndrome: Presence of bipolar disorder and progressive loss of skills. Mol Syndromol 3:14–20.

Du Y, Weed SA, Xiong WC, Marshall TD, Parsons JT. 1998. Identification of a novel cortactin SH3 domain-binding protein and its localization to growth cones of cultured neurons. Mol Cell Biol 18:5838–5851.

Durand CM, Betancur C, Boeckers TM, Bockmann J, Chaste P, Fauchereau F, Nygren G, et al. 2007. Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. Nat Genet 39:25–27.

Ey E, Yang M, Katz AM, Woldeyohannes L, Silverman JL, Leblond CS, Faure P, et al. 2012. Absence of deficits in social behaviors and ultrasonic vocalizations in later generations of mice lacking neuroligin4. Genes Brain Behav 11:928–941.

Failla P, Romano C, Alberti A, Vasta A, Buono S, Castiglia L, Luciano D, et al. 2007. Schizophrenia in a patient with subtelomeric duplication of chromosome 22q. Clin Genet 71:599–601.

Flint J, Wilkie AO, Buckle VJ, Winter RM, Holland AJ, McDermid HE. 1995. The detection of subtelomeric chromosomal rearrangements in idiopathic mental retardation. Nat Genet 9:132–140.

Gauthier J, Champagne N, Lafreniere RG, Xiong L, Spiegelman D, Brustein E, Lapointe M, et al. 2010. De novo mutations in the gene encoding the synaptic scaffolding protein SHANK3 in patients ascertained for schizophrenia. Proc Natl Acad Sci USA 107:7863–7868.

Gauthier J, Spiegelman D, Piton A, Lafreniere RG, Laurent S, St-Onge J, Lapointe L, et al. 2009. Novel de novo SHANK3 mutation in autistic patients. Am J Med Genet B Neuropsychiatr Genet 150B:421–424.

Gong Y, Lippa CF, Zhu J, Lin Q, Rosso AL. 2009. Disruption of glutamate receptors at Shank-postsynaptic platform in Alzheimer's disease. Brain Res 1292: 191–198.

Grabrucker AM, Schmeisser MJ, Schoen M, Boeckers TM. 2011. Postsynaptic ProSAP/Shank scaffolds in the cross-hair of synaptopathies. Trends Cell Biol 21:594–603.

Hardy J, Selkoe DJ. 2002. The amyloid hypothesis of Alzheimer's disease: Progress and problems on the road to therapeutics. Science 297:353–356.

Hung AY, Futai K, Sala C, Valtschanoff JG, Ryu J, Woodworth MA, Kidd FL, et al. 2008. Smaller dendritic spines, weaker synaptic transmission, but enhanced spatial learning in mice lacking Shank1. J Neurosci 28: 1697–1708.

Hutton M, Lendon CL, Rizzu P, Baker M, Froelich S, Houlden H, Pickering-Brown S, et al. 1998. Association of missense and 5′-splice-site mutations in tau with the inherited dementia FTDP-17. Nature 393:702–705.

Jamain S, Quach H, Betancur C, Rastam M, Colineaux C, Gillberg IC, Soderstrom H, et al. 2003. Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. Nat Genet 34:27–29.

Leblond CS, Heinrich J, Delorme R, Proepper C, Betancur C, Huguet G, Konyukh M, et al. 2012. Genetic and functional analyses of SHANK2 mutations suggest a multiple hit model of autism spectrum disorders. PLoS Genet 8: e1002521.

Lennertz L, Wagner M, Wolwer W, Schuhmacher A, Frommann I, Berning J, Schulze-Rauschenbach S, et al. 2012. A promoter variant of SHANK1 affects auditory working memory in schizophrenia patients and in subjects clinically at risk for psychosis. Eur Arch Psychiatry Clin Neurosci 262:117–124.

Lesca G, Rudolf G, Labalme A, Hirsch E, Arzimanoglou A, Genton P, Motte J, et al. 2012. Epileptic encephalopathies of the Landau-Kleffner and continuous spike and waves during slow-wave sleep types: Genomic dissection makes the link with autism. Epilepsia 53: 1526–1538.

Lim S, Naisbitt S, Yoon J, Hwang JI, Suh PG, Sheng M, Kim E. 1999. Characterization of the Shank family of synaptic proteins. Multiple genes, alternative splicing, and differential expression in brain and development. J Biol Chem 274:29510–29518.

Misceo D, Rodningen OK, Baroy T, Sorte H, Mellembakken JR, Stromme P, Fannemel M, et al. 2011. A translocation between Xq21.33 and 22q13.33 causes an intragenic SHANK3 deletion in a woman with Phelan-McDermid syndrome and hypergonadotropic hypogonadism. Am J Med Genet A 155A:403–408.

Moessner R, Marshall CR, Sutcliffe JS, Skaug J, Pinto D, Vincent J, Zwaigenbaum L, et al. 2007. Contribution of SHANK3 mutations to autism spectrum disorder. Am J Hum Genet 81:1289–1297.

Naisbitt S, Kim E, Tu JC, Xiao B, Sala C, Valtschanoff J, Weinberg RJ, et al. 1999. Shank, a novel family of postsynaptic density proteins that binds to the NMDA receptor/PSD-95/GKAP complex and cortactin. Neuron 23: 569–582.

Nesslinger NJ, Gorski JL, Kurczynski TW, Shapira SK, Siegel-Bartelt J, Dumanski JP, Cullen RF Jr., et al. 1994. Clinical, cytogenetic, and molecular characterization of seven patients with deletions of chromosome 22q13.3. Am J Hum Genet 54:464–472.

Peca J, Feliciano C, Ting JT, Wang W, Wells MF, Venkatraman TN, Lascola CD, et al. 2011. Shank3 mutant mice display autistic-like behaviours and striatal dysfunction. Nature 472:437–442.

Pham E, Crews L, Ubhi K, Hansen L, Adame A, Cartier A, Salmon D, et al. 2010. Progressive accumulation of amyloid-beta oligomers in Alzheimer's disease and in amyloid precursor protein transgenic mice is accompanied by selective alterations in synaptic scaffold proteins. FEBS J 277:3051–3067.

Phelan K, McDermid HE. 2012. The 22q13.3 deletion syndrome (Phelan-McDermid syndrome). Mol Syndromol 2:186–201.

Phelan MC, Rogers RC, Saul RA, Stapleton GA, Sweet K, McDermid H, Shaw SR, et al. 2001. 22q13 deletion syndrome. Am J Med Genet 101:91–99.

Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R, Conroy J, et al. 2010. Functional impact of global rare copy number variation in autism spectrum disorders. Nature 466:368–372.

Quon D, Wang Y, Catalano R, Scardina JM, Murakami K, Cordell B. 1991. Formation of beta-amyloid protein deposits in brains of transgenic mice. Nature 352:239–241.

Roselli F, Hutzler P, Wegerich Y, Livrea P, Almeida OF. 2009. Disassembly of shank and homer synaptic clusters is driven by soluble beta-amyloid(1-40) through divergent NMDAR-dependent signalling pathways. PLoS One 4:e6011.

Rovelet-Lecrux A, Hannequin D, Guillin O, Legallic S, Jurici S, Wallon D, Frebourg T, et al. 2010. Frontotemporal dementia phenotype associated with MAPT gene duplication. J Alzheimers Dis 21:897–902.

Sato D, Lionel AC, Leblond CS, Prasad A, Pinto D, Walker S, O'Connor I, et al. 2012. SHANK1 deletions in males with autism spectrum disorder. Am J Hum Genet 90: 879–887.

Schluth-Bolard C, Labalme A, Cordier MP, Till M, Nadeau G, Tevissen H, Lesca G, et al. 2013. Breakpoint mapping by next generation sequencing reveals causative gene disruption in patients carrying apparently balanced chromosome rearrangements with intellectual deficiency and/or congenital malformations. J Med Genet 50:144–150.

Schmeisser MJ, Ey E, Wegener S, Bockmann J, Stempel AV, Kuebler A, Janssen AL, et al. 2012. Autistic-like behaviours and hyperactivity in mice lacking ProSAP1/Shank2. Nature 486:256–260.

Shaw-Smith C, Pittman AM, Willatt L, Martin H, Rickman L, Gribble S, Curley R, et al. 2006. Microdeletion encompassing MAPT at chromosome 17q21.3 is associated with developmental delay and learning disability. Nat Genet 38:1032–1037.

Silverman JL, Turner SM, Barkan CL, Tolu SS, Saxena R, Hung AY, Sheng M, et al. 2011. Sociability and motor functions in Shank1 mutant mice. Brain Res 1380: 120–137.

Sovner R, Stone A, Fox C. 1996. Ring chromosome 22 and mood disorders. J Intellect Disabil Res 40 (Part 1):82–86.

Sudhof TC. 2008. Neuroligins and neurexins link synaptic function to cognitive disease. Nature 455:903–911.

Toro R, Konyukh M, Delorme R, Leblond C, Chaste P, Fauchereau F, Coleman M, et al. 2010. Key role for gene dosage and synaptic homeostasis in autism spectrum disorders. Trends Genet 26:363–372.

Verhoeven WM, Egger JI, Cohen-Snuijf R, Kant SG, de Leeuw N. 2013. Phelan-McDermid syndrome: Clinical report of a 70-year-old woman. Am J Med Genet A 161A:158–161.

Verhoeven WM, Egger JI, Willemsen MH, de Leijer GJ, Kleefstra T. 2012. Phelan-McDermid syndrome in two adult brothers: Atypical bipolar disorder as its psychopathological phenotype? Neuropsychiatr Dis Treat 8: 175–179.

Verpelli C, Schmeisser MJ, Sala C, Boeckers TM. 2012. Scaffold proteins at the postsynaptic density. Adv Exp Med Biol 970:29–61.

Vucurovic K, Landais E, Delahaigue C, Eutrope J, Schneider A, Leroy C, Kabbaj H, et al. 2012. Bipolar affective disorder and early dementia onset in a male patient with SHANK3 deletion. Eur J Med Genet 55: 625–629.

Waga C, Okamoto N, Ondo Y, Fukumura-Kato R, Goto Y, Kohsaka S, Uchino S. 2011. Novel variants of the SHANK3 gene in Japanese autistic patients with severe delayed speech development. Psychiatr Genet 21: 208–211.

Wang X, McCoy PA, Rodriguiz RM, Pan Y, Je HS, Roberts AC, Kim CJ, et al. 2011. Synaptic dysfunction and abnormal behaviors in mice lacking major isoforms of Shank3. Hum Mol Genet 20: 3093–3108.

Watt JL, Olson IA, Johnston AW, Ross HS, Couzin DA, Stephen GS. 1985. A familial pericentric inversion of chromosome 22 with a recombinant subject illustrating a 'pure' partial monosomy syndrome. J Med Genet 22: 283–287.

Willemsen MH, Rensen JH, van Schrojenstein-Lantman de Valk HM, Hamel BC, Kleefstra T. 2012. Adult phenotypes in Angelman- and Rett-like syndromes. Mol Syndromol 2:217–234.

Wilson HL, Crolla JA, Walker D, Artifoni L, Dallapiccola B, Takano T, Vasudevan P, et al. 2008. Interstitial 22q13 deletions: Genes other than SHANK3 have major effects on cognitive and language development. Eur J Hum Genet 16:1301–1310.

Wischmeijer A, Magini P, Giorda R, Gnoli M, Ciccone R, Cecconi L, Franzoni E, et al. 2011. Olfactory receptor-related duplicons mediate a microdeletion at 11q13.2q13.4 associated with a syndromic phenotype. Mol Syndromol 1: 176–184.

Wohr M, Roullet FI, Hung AY, Sheng M, Crawley JN. 2011. Communication impairments in mice lacking Shank1: Reduced levels of ultrasonic vocalizations and scent marking behavior. PLoS One 6:e20631.

Won H, Lee HR, Gee HY, Mah W, Kim JI, Lee J, Ha S, et al. 2012. Autistic-like social behaviour in Shank2-mutant mice improved by restoring NMDA receptor function. Nature 486:261–265.

Yang M, Bozdagi O, Scattoni ML, Wohr M, Roullet FI, Katz AM, Abrams DN, et al. 2012. Reduced excitatory neurotransmission and mild autism-relevant phenotypes in adolescent Shank3 null mutant mice. J Neurosci 32: 6525–6541.

The Genetic Landscapes of Autism Spectrum Disorders

Guillaume Huguet,[1,2,3] Elodie Ey,[1,2,3] and Thomas Bourgeron[1,2,3]

[1]Human Genetics and Cognitive Functions Unit, Institut Pasteur, 75015 Paris, France; email: thomasb@pasteur.fr

[2]CNRS URA 2182 (Genes, Synapses, and Cognition), Institut Pasteur, 75015 Paris, France

[3]Human Genetics and Cognitive Functions Unit, Université Paris Diderot, Sorbonne Paris Cité, 75013 Paris, France

Annu. Rev. Genomics Hum. Genet. 2013. 14:191–213

First published online as a Review in Advance on July 22, 2013

The *Annual Review of Genomics and Human Genetics* is online at genom.annualreviews.org

This article's doi: 10.1146/annurev-genom-091212-153431

Keywords

copy-number variant, synapse, de novo mutation, animal models

Abstract

The autism spectrum disorders (ASD) are characterized by impairments in social interaction and stereotyped behaviors. For the majority of individuals with ASD, the causes of the disorder remain unknown; however, in up to 25% of cases, a genetic cause can be identified. Chromosomal rearrangements as well as rare and de novo copy-number variants are present in ~10–20% of individuals with ASD, compared with 1–2% in the general population and/or unaffected siblings. Rare and de novo coding-sequence mutations affecting neuronal genes have also been identified in ~5–10% of individuals with ASD. Common variants such as single-nucleotide polymorphisms seem to contribute to ASD susceptibility, but, taken individually, their effects appear to be small. Despite a heterogeneous genetic landscape, the genes implicated thus far—which are involved in chromatin remodeling, metabolism, mRNA translation, and synaptic function—seem to converge in common pathways affecting neuronal and synaptic homeostasis. Animal models developed to study these genes should lead to a better understanding of the diversity of the genetic landscapes of ASD.

THE CLINICAL AND GENETIC HETEROGENEITY OF ASD

The autism spectrum disorders (ASD) are early-onset neuropsychiatric disorders characterized by impaired social interactions, restricted interests, and repetitive behaviors (7, 63). They are present in 0.7–1.1% of the population and are four times more common in males than in females (39). The core symptoms of ASD rarely come in isolation, typically coexisting with other psychiatric and medical conditions such as intellectual disability, epilepsy, motor control problems, attention-deficit/hyperactivity disorder (ADHD), tics, anxiety, sleep disorders, and gastrointestinal problems. A common feature in ASD is an abnormal (too high or too low) response to sensory stimuli. In some cases, abnormality in auditory, visual, and somatosensory information processing might even be the primary factor driving abnormal development of socialization and communication skills (73). The concept of early symptomatic syndromes eliciting neurodevelopmental clinical examinations (ESSENCE) was recently introduced to better take into account the clinical heterogeneity and syndromic overlap of impairing developmental symptoms in children under the ages of 3–5 (46). Major problems in at least one ESSENCE domain in the fields of general development, communication and language, social interrelatedness, motor coordination, attention, activity, behavior, mood, and/or sleep before age 5 often signal major problems in the same or overlapping domains years later (46).

There is a strong genetic component to ASD susceptibility, as indicated by the recurrence risk in families and twin studies (1, 34, 44). The recurrence risk in families with one child with ASD was initially estimated at 5%, compared with 1% in the general population, but recent studies investigating early signs of ASD in siblings of individuals with ASD revealed an even higher recurrence rate (up to 20%) (91). The first twin studies described ASD as the most genetic of neuropsychiatric disorders, with concordance rates of 82–92% in monozygotic twins compared with 1–10% in dizygotic twins (9, 43); more recent studies, however, have indicated that the concordance in dizygotic twins might be higher (>20%) (53). In all studies, the concordance in monozygotic twins is not complete, indicating that epigenetic, stochastic, and/or environmental factors are present (75, 77, 98).

As illustrated in the epigenetic landscape of Conrad Hal Waddington (126), the ASD phenotype may result from a complex regulatory network involving genetic, epigenetic, and environmental factors as well as stochastic fluctuations, much as a marble is guided toward the point of lowest local elevation (**Figure 1*b***). For each individual with a given genotype, the response to the environment might be different. Woltereck (130) introduced the concepts of norms of reaction and phenotypic plasticity at the beginning of the twentieth century to account for the fact that organisms can change their characteristics in response to the environment (**Figure 1*c***). In line with this idea, different genotypes that increase the risk of ASD could have distinct norms of reaction across a range of environments. Thus, any information on the genetic causes of ASD should help us identify the best environment(s) for a given individual.

The first genetic causes of ASD were identified in monogenic disorders such as fragile X and Rett syndromes (3, 95) and in families carrying chromosomal abnormalities (123). Later, the identification of deleterious mutations in several candidate genes, such as those encoding neuroligins, neurexins, and SHANKs, pointed at the synapse as a key player in ASD susceptibility, including in individuals with an ASD but no intellectual deficiency, such as those with Asperger syndrome (36, 62, 114). Whole-genome microarray studies then revealed submicroscopic deletions and duplications, called copy-number variants, that affect many loci, including de novo events in 5–15% of individuals with ASD (26, 29, 47, 50, 59, 80, 96, 100, 105, 114). More recently, exome sequencing has been used to detect deleterious de novo mutations in 3.6–8.8% of individuals with ASD (68, 86, 88, 90, 101).

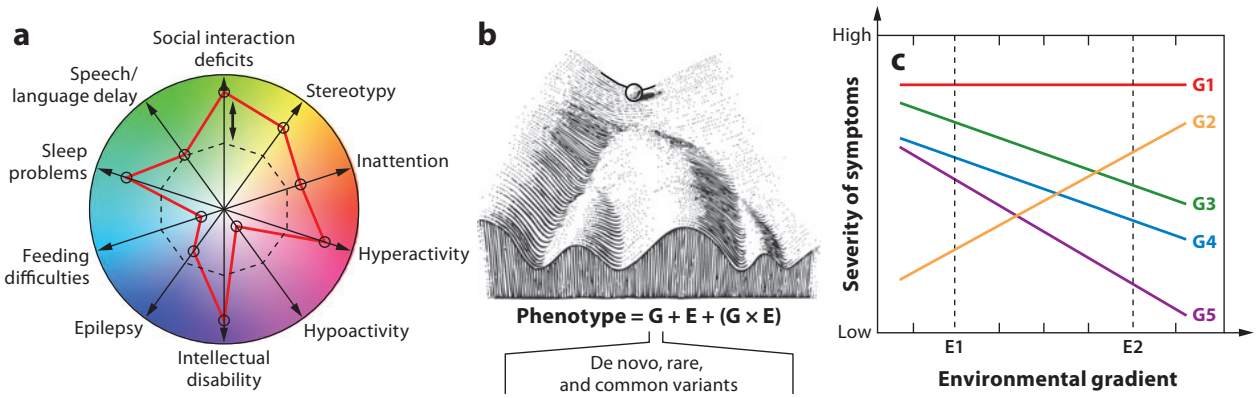


Figure 1

Clinical and genetic diversity of ASD. (*a*) Core symptoms and frequent comorbidities observed in individuals with ASD. Arrows define clinical characteristics of the individuals, and the red line defines the level of severity (low at the center and high at the periphery). Colors represent the combination of clinical features. (*b*) An epigenetic landscape. As with a marble guided toward the lowest local point, the clinical outcome of an individual with ASD is influenced by genetic factors (G), environmental factors (E), and gene-environment interactions (G × E). Adapted from Waddington (126). (*c*) Norms of reaction describing how individuals with different genotypes respond to different environments. The individual with genotype G1 does not improve regardless of the environment. The individual with genotype G2 worsens in environment E2 compared with environment E1. Individuals with genotypes G3–G5 improve in environment E2 compared with environment E1, but with different norms of reaction. Adapted from Woltereck (130).

In summary, beyond the unifying definition of ASD lies an extreme degree of clinical heterogeneity, ranging from profound to moderate impairments. Molecular genetic studies have revealed the genetic landscape of ASDs to be highly heterogeneous, with different types of genetic abnormalities located on almost all chromosomes with varying levels of penetrance (1, 34, 44).

DE NOVO COPY-NUMBER VARIANTS IN ASD

Microscopically visible chromosomal alterations have been reported in fewer than 5% of individuals with ASD. The most frequent are located at chromosomal regions 2q37, 7q11, 15q11–13, 16p11, 22q11.2, and 22q13.3 (123). Over the past decade, important advances in the genetics of ASD have emerged from studies of copy-number variants using whole-genome microarrays (2). The first analysis, which used comparative genomic hybridization (CGH) on a small sample of 29 individuals with ASD, found that 8 of them (28%) had clinically relevant copy-number variants (six deletions and two duplications) (59). The high rate of genetic alterations reported in that study was most likely due to the inclusion of individuals who displayed other comorbidities commonly associated with genetic abnormalities, such as facial dysmorphism. The second main study used a denser CGH array and was not biased toward individuals with comorbidities; here, the authors found de novo copy-number variants in 12 of 118 individuals with sporadic ASD (10%), 2 of 77 individuals with ASD who also had an affected first-degree relative (3%), and 2 of 196 controls (1%) (105).

Numerous studies have since used CGH or single-nucleotide polymorphism (SNP) arrays to investigate the burden of copy-number variants in large cohorts of individuals with ASD (26, 29, 47, 50, 58, 59, 80, 96, 100, 105, 114). To date, more than 3,800 individuals with ASD, 1,200 unaffected siblings, and 600 controls along with the genotypes of the parents have been analyzed. In all studies, de novo copy-number variants appeared to be enriched in individuals with ASD compared with their unaffected siblings or the control population. Overall, 6.6% of individuals

with sporadic ASD had at least one rare de novo copy-number variant, compared with 4.1% in individuals with ASD who also had an affected first-degree relative, 1.4% in unaffected siblings, and 1.9% in controls (**Figure 2*a***). Notably, the de novo copy-number variants are larger and affect more genes in the individuals with ASD compared with their unaffected siblings and controls. This enrichment does not reflect a higher risk of parents to produce these de novo variants, because the unaffected siblings display a frequency of de novo variants similar to controls. As expected, de novo copy-number variants are more frequent in individuals with sporadic ASD than in individuals with ASD who also had an affected first-degree relative, indicating that the structure of the family needs to be considered to estimate the prevalence of the de novo copy-number variants in ASD.

Two large-scale studies determined the parental origin of the de novo copy-number variants (58, 100). Combining the results of the two studies reveals a trend of excess maternal de novo copy-number variants (54.5%), but the pattern does not reach statistical significance (**Figure 2*b***). Interestingly, compared with males, females with ASD have more de novo copy-number variants, and their genomic imbalances are larger and affect more genes (**Figure 2*c***).

Finally, there is a significant association between the cognitive level as measured by IQ and the number of genes affected by de novo copy-number variants (48, 100). In a study by Girirajan et al. (48), the median IQ crossed the threshold for intellectual disability (IQ < 70) in individuals with 18 or more disrupted genes (**Figure 2*d***).

In summary, the burden of rare and de novo copy-number variants is significantly higher in individuals with ASD compared with their unaffected siblings and the general population. The variants are more numerous, larger, and contain more genes. The presence of dysmorphic features and/or intellectual disability increases the chances of detecting a de novo copy-number variant. Finally, females seem to have a greater tolerance than males for aggregating genetic alterations without developing ASD. Yet if copy-number variants are strongly associated with ASD, especially in sporadic cases, then they account for only a small fraction of individuals with ASD, and therefore other types of mutations must play a role in the susceptibility to ASD, especially in familial cases.

DE NOVO CODING-SEQUENCE VARIANTS IN ASD

Investigations of de novo coding-sequence variants were previously limited to specific candidate genes. Next-generation sequencing technologies, such as whole-exome and whole-genome sequencing, now allow us to estimate the contribution of de novo coding-sequence mutations to ASD. The six exome studies published to date have sequenced more than 1,000 individuals with ASD along with 790 unaffected siblings (57, 68, 86, 88, 90, 101). In these studies, the average rate of de novo coding-sequence variants per individual (including missense, splicing, frameshift, and stop-gain variants) was 0.86% in females with ASD, 0.73% in males with ASD, and 0.60% in unaffected male and female siblings (**Figure 3*a***). At the group level, there was no statistically significant difference between individuals with ASD, unaffected siblings, and controls. Only when the analysis is restricted to genes expressed in the brain does the difference reach statistical significance ($p = 0.001$), with more mutations in individuals with ASD compared with unaffected siblings (101). At the individual level, several causative mutations have been identified, accounting for 3.6–8.8% of individuals with ASD. Iossifov et al. (57) have estimated that 65 causative genes have been identified by four large-scale exome sequencing studies. This calculation is based on the number of deleterious de novo mutations identified in the different exome sequencing studies (59 of 343 probands in Reference 57, 90 of 175 probands in Reference 86, 33 of 209 probands in Reference 90, and 17 of 238 probands in Reference 100) and on an estimated twofold differential rate of deleterious mutations between individuals with ASD and their siblings. As for copy-number

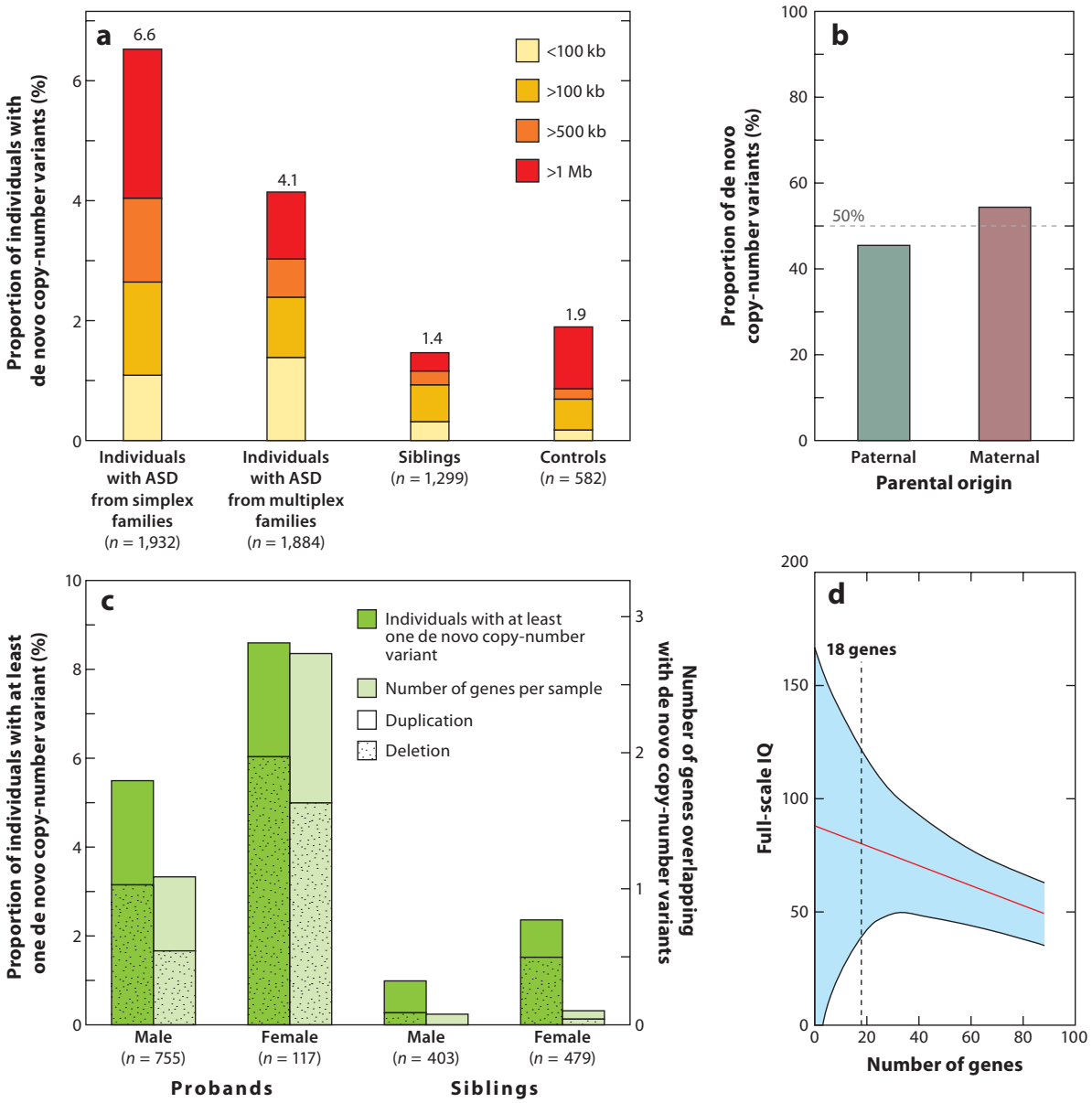


Figure 2

The burden of de novo copy-number variants in ASD. (*a*) Frequency and size distributions of de novo copy-number variants in individuals with ASD from simplex families, individuals with ASD from multiplex families, unaffected siblings, and controls. Data are from References 58, 80, 96, 100, and 105. (*b*) Parental origin of de novo copy-number variants identified in individuals with ASD. Data are from References 58 and 100. (*c*) Proportion of individuals with at least one de novo copy-number variant, along with the number of RefSeq genes altered by genomic imbalance. Data are from Reference 100. (*d*) Relationship between full-scale IQ and the number of genes overlapping with de novo copy-number variants. Data are from References 48 and 100.

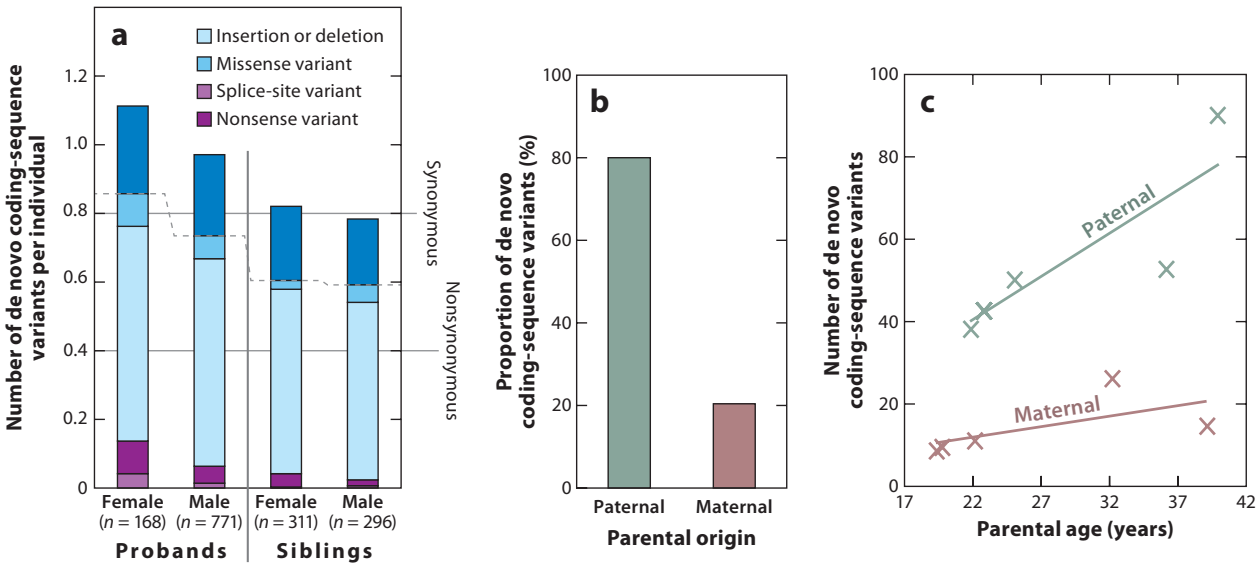


Figure 3

The burden of de novo coding-sequence variants in ASD. (*a*) Distribution of de novo coding-sequence variants in individuals with ASD and controls after stratification by sex. Data are from References 57, 86, 90, and 100. (*b*) Parental origin of de novo coding-sequence variants identified in individuals with ASD. Data are from References 68 and 90. (*c*) Relationship between parental age and the number of de novo coding-sequence variants in the child. Data are from Reference 68.

variants in individuals with ASD, females tend to have more de novo coding-sequence variants compared with males. When the parental origin of the de novo mutations could be ascertained, variants were three times more likely to come from the paternal chromosome than from the maternal one (68, 90) (**Figure 3*b***). The age of the father at conception greatly influenced this rate (68), with an increase of almost two mutations per year. An exponential model estimates that the number of paternal mutations doubles every 16.5 years (**Figure 3*c***).

One disappointing result originating from exome sequencing studies was that, with the exception of a handful of genes (*CHD8*, *KATNAL2*, *SCN1A*, *SCN2A*, *DYRK1A*, and *POGZ*), most mutations were restricted to a single individual (57, 86, 90, 101). The absence of major genes for ASD—that is, genes that could account for more than 1% of the cases—was already suggested by the lack of strong linkage and association signals. Nevertheless, it is important to keep in mind that the current methodologies used for exome sequencing do not capture all exons of the genome (especially when they are in GC-rich sequences) and that new sequencing technologies might be able to detect additional mutations (89). Stratifying individuals with ASD based on intermediate phenotypes or symptoms might also detect genes with a high prevalence of mutation in a subgroup of individuals with ASD, such as *CHD8*, which is recurrently mutated in individuals with macrocephaly (89), and *SHANK3*, which is mutated in up to 2–3% of individuals with ASD and intellectual deficiency (C.S. Leblond, C. Nava, A. Polge, J. Gauthier, G. Huguet, et al., unpublished data).

In summary, the burden of de novo mutations affecting genes expressed in the brain is higher in individuals with ASD compared with controls. Two-thirds of the de novo mutations are of paternal origin, and the rate of these mutations increases with age. None of the identified genes account for more than 1% of the individuals with ASD, and the current predictions estimate that more than 500–1,000 genes could be associated with ASD (57, 101).

RECURRENT DE NOVO AND INHERITED COPY-NUMBER VARIANTS IN ASD

Among the copy-number variants recurrently identified in individuals with ASD, many are associated with clinically well-defined genetic syndromes, such as Smith-Magenis syndrome, Williams-Beuren syndrome, Sotos syndrome, Prader-Willi and Angelman syndromes, 15q13.3 microdeletion/duplication syndrome, and 17q21.31 deletion syndrome (123). The majority of these loci are flanked by segmental duplications, which greatly increase the risk of deletion or duplication events. Remarkably, these copy-number variants can be either de novo or inherited from unaffected parents. Three loci on chromosomal regions 7q11, 15q11.2–13.3, and 16p11.2 have been strongly associated with ASD (11, 70, 100, 113, 129). For each of these loci, the variant was inherited in at least one instance, and the extent to which the loci were inherited (versus de novo occurrence) ranged from 5.6% to 100% (48). One striking property of these loci is that they are associated with highly contrasting clinical outcomes depending on whether the region was deleted or duplicated.

On chromosomal region 7q11.23 (**Figure 4*a***), deletions are associated with Williams-Beuren syndrome, a rare neurodevelopmental disorder affecting 1 out of 10,000 newborns that is characterized by facial dysmorphism, short stature, intellectual disability, and a specific cognitive profile with strengths in verbal short-term memory and language but strong difficulties in visuospatial tasks. The same chromosomal region, when duplicated, is associated with ASD, ADHD, intellectual disability, language impairment, and epilepsy (31, 100, 110). In one study, the deletions were mostly de novo, but two-thirds of the duplications were inherited (118). The critical region of 1.5–1.8 Mb contains approximately 30 genes. The haploinsufficiency of the elastin-encoding gene is responsible for the cardiovascular problems of individuals with Williams-Beuren syndrome, but the genes causing the cognitive defects remain to be identified. Several genes expressed in the brain are compelling candidates (e.g., *TRIM50*, *FKBP6*, *FZD9*, *STX1A*, *LIMK1*, and *EIF4H*). Among these genes, three encode proteins that directly participate in synaptic function (**Figure 5**). STX1A forms a complex with SNAP25 and VAMP2 (also known as synaptobrevin) and is involved in synaptic vesicle exocytosis. *LIMK1* codes for a serine protein kinase and regulates dendritic spine morphology by remodeling the actin skeleton. EIF4H is a translation initiation factor that regulates translation at the synapse together with the fragile X protein FMRP, the elongation factor EIF4E, and CYFIP1, three proteins that are mutated in individuals with ASD (87) (**Figure 5**).

On chromosomal region 15q11.1–13.3 (**Figure 4*b***), a cluster of low-copy repeats greatly increases the risk of complex genomic rearrangements at five distinct breakpoints (low-copy repeat sequences), BP1–BP5 (113). Paternal deletions between BP1 and BP3 (type I) and BP2 and BP3 (type II) cause Prader-Willi syndrome in 1 out of 15,000–30,000 live births, whereas maternal deletions of the same region cause Angelman syndrome in 1 out of 12,000–20,000 live births (78). Individuals with these syndromes can also fulfill a diagnosis of ASD. In a survey of 150 individuals with Prader-Willi syndrome and 104 individuals with Angelman syndrome, the rate of ASD was 25.3% in the former group and 1.9% in the latter (120). Maternal duplications of the 15q11.1–13.3 region are one of the major genetic causes of ASD, accounting for approximately 1% of individuals with ASD (28, 29, 32). Several genes in the BP2–BP3 deleted region are subject to parental imprinting, and uniparental disomy of this region of chromosome 15 is therefore also a cause of Prader-Willi and Angelman syndromes. The *UBE3A* gene is responsible for Angelman syndrome (66, 81, 112) and codes for a ubiquitin ligase that is expressed exclusively from the maternal alleles in neurons (132) (**Figure 5**).

Interestingly, individuals carrying deletions between BP1 and BP3 usually have a more severe phenotype than individuals with deletions between BP2 and BP3 (20, 54). In the small

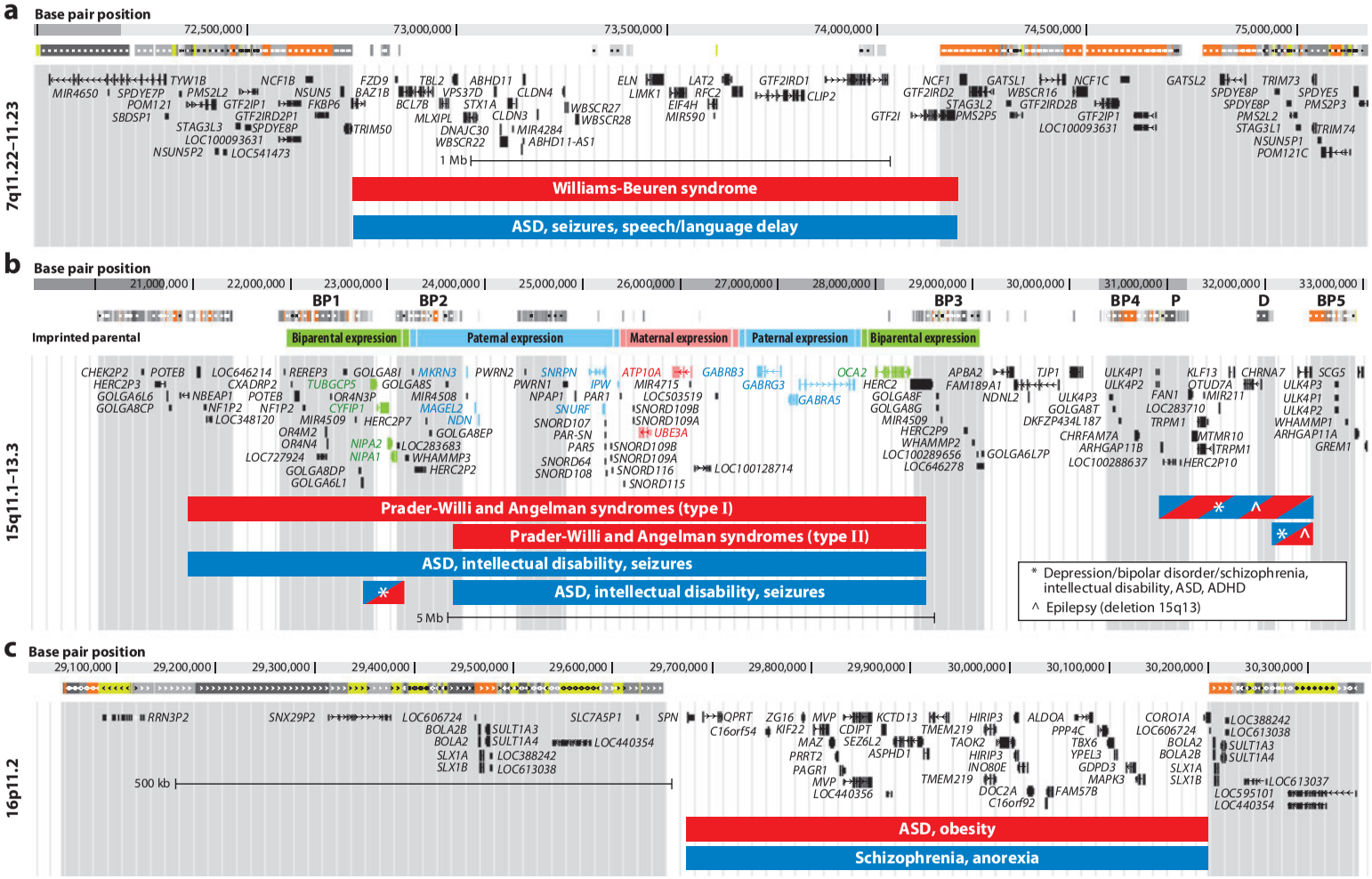


Figure 4

Recurrent rare and de novo copy-number variants in ASD. (*a*). On chromosomal region 7q11, deletions (*red*) are associated with Williams-Beuren syndrome, and duplications (*blue*) are associated with ASD, seizures, and speech/language delay. (*b*) On chromosomal region 15q11, deletions and duplications are located at different breakpoints (BP1–5, P, and D) made of low-copy repeat sequences. The maternal deletions are associated with Angelman syndrome; the paternal deletions are associated with Prader-Willi syndrome. The maternal duplications are associated with ASD, and both maternal and paternal duplications are associated with intellectual disability and seizures. (*c*) On chromosomal region 16p11.2, deletions are associated with ASD, macrocephaly, and obesity, and duplications are associated with schizophrenia, microcephaly, and anorexia. Abbreviations: BP1–5, breakpoints 1–5; P and D, proximal and distal low-copy repeats, respectively.

BP1–BP2 region (500 kb), four genes are involved in brain function: *CYFIP1*, *NIPA1*, *NIPA2*, and *TUBGCP5*. *CYFIP1* encodes a protein that interacts with FMRP, the protein responsible for fragile X syndrome, to repress synaptic translation (84, 103). *NIPA2* was recently identified as encoding a magnesium transporter, and *TUBGCP5* encodes a member of the cytoskeleton tubulin complex. Both deletions and duplications that are restricted to BP1–BP2 are associated with delays in cognitive and speech development, ASD, ADHD, and schizophrenia (17, 119). However, the BP1–BP2 copy-number variants are not fully penetrant; they are observed in 0.8% of individuals with ASD (0.4% deletions and 0.4% duplications) and in 0.38% of the general population (0.25% deletions and 0.13% duplications) (19).

Finally, the BP4–BP5 region contains six genes (*MTMR15*, *MTMR10*, *TRPM1*, *KLF13*, *OTUD7A*, and *CHRNA7*) and one microRNA (hsamir211). In this interval, a complex set of internal rearrangements is observed owing to the presence of additional proximal and distal low-copy repeat sequences (P and D, respectively). The deletions of *CHRNA7* coding for the nicotinic $\alpha 7$ receptors are associated with a broad range of developmental delays, intellectual disability, depression, bipolar disorder, ASD, ADHD, and epilepsy (14, 35, 55, 56, 82, 92, 106, 107, 111, 117). Duplications of *CHRNA7* are also associated with a risk of cognitive deficits and neurobehavioral abnormalities, but with an apparently lower penetrance compared with the deletions (35, 55, 107).

On chromosomal region 16p11.2 (**Figure 4*c***), several segmental duplications increase the risk of genomic rearrangements. A recurrent deletion of a ~600-kb region containing 29 genes (BP4–BP5) is observed in 1 out of 2,000 individuals from the general population and reaches 0.5% in ASD (11, 12, 16, 42, 49, 70, 129). There is a shift in the IQ distribution for carriers of this deletion (mean verbal IQ = 74 and mean nonverbal IQ = 83), and a majority of these individuals require speech therapy (61). More than 80% of individuals carrying the deletion exhibit psychiatric disorders, including ASD, which is present in 15% of the pediatric carriers (61). Interestingly, the deletion is associated with obesity and macrocephaly, whereas the duplication is associated with anorexia and microcephaly (61). Obesity is a major comorbidity present in 50% of the deletion carriers by the age of 7, but it does not correlate with IQ or any other behavioral trait (61). Seizures are present in 24% of carriers and occur independently of other symptoms. Among the 29 genes, 17 are compelling candidates for ASD (69, 71). Overexpression of each of the 17 human transcripts in zebrafish embryos identified *KCTD13* as the sole gene capable of inducing microcephaly when the transcript was overexpressed (similar to a 16p11.2 duplication) and macrocephaly when it was suppressed (similar to a 16p11.2 deletion) (51). *KCTD13* encodes PDIP1, which interacts with PCNA27 and thus might have a role in regulating the cell cycle during neurogenesis.

In summary, recurrent chromosomal rearrangements are associated with ASD. The striking observation that deletions and duplications lead to different clinical outcomes indicates that a fine balance of correct gene dosage is necessary for human brain function and development. Further studies are warranted to identify the causative genes (or, most likely, the combination of causative genes) that increase the risk of ASD in individuals carrying these genomic rearrangements.

MULTIPLE HITS IN ASD

Given that de novo mutations account for only a small fraction of individuals with ASD (<15%), inherited variants are expected to play a major role in genetic susceptibility to ASD. In addition, inherited risk and protective alleles might contribute to the incomplete penetrance and variable expressivity observed in individuals carrying de novo deleterious mutations. Several studies have demonstrated the presence of more than one deleterious mutation (multiple hits) in individuals with ASD (48, 49, 74). In a large-scale study of 2,312 children known to carry a copy-number variant associated with intellectual disability and congenital abnormalities, 10% carried a second

large copy-number variant in addition to the primary genetic lesion (48). Children who carried two large copy-number variants of unknown clinical significance were eight times more likely than controls to have developmental delays. Among affected children, inherited copy-number variants tended to co-occur with a second-site large copy-number variant. No parental bias was observed for the primary de novo or inherited site, but 72% of the second-site copy-number variants were inherited from the mother (48).

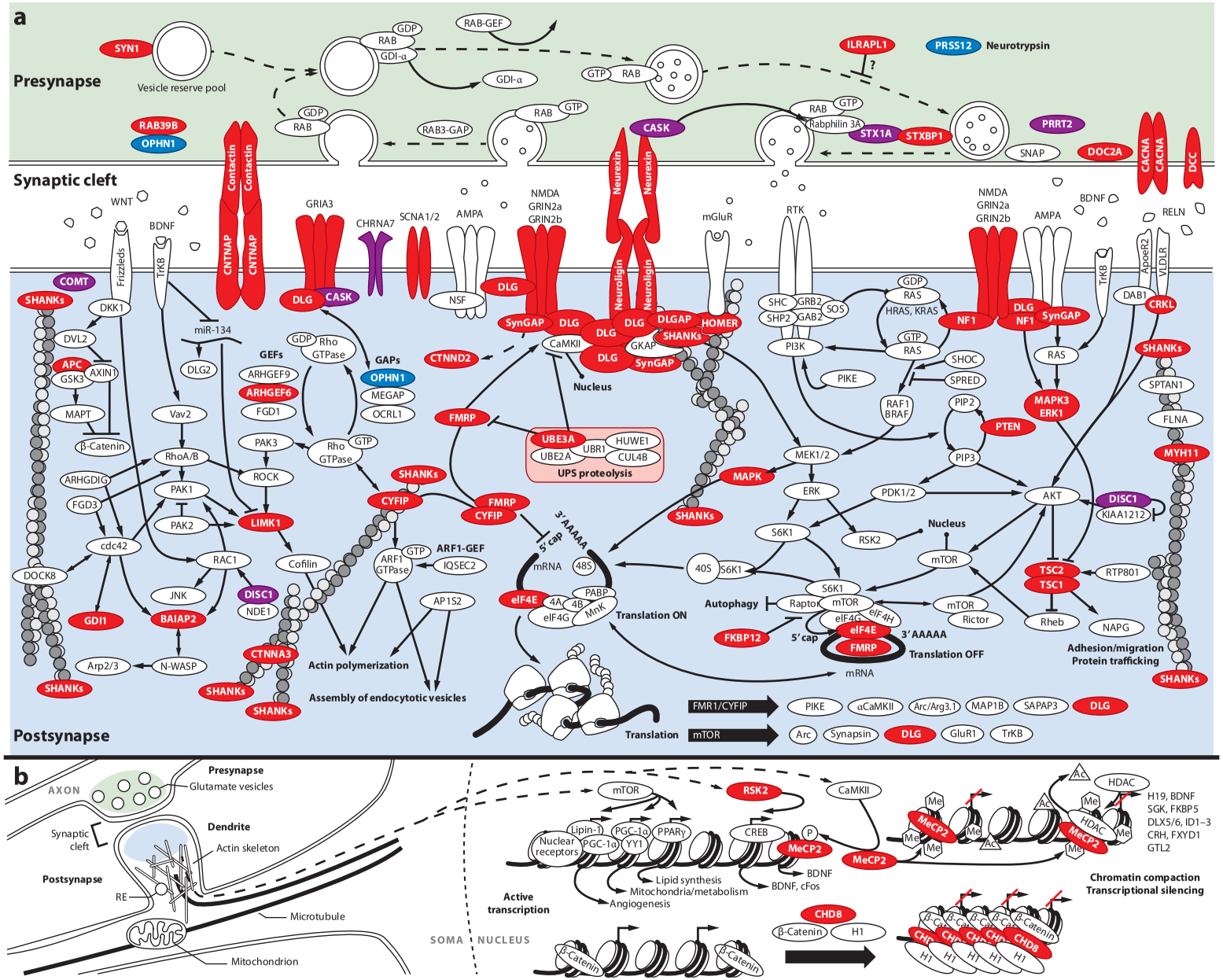
Other studies have supported a multiple-hits model. In 42 carriers of a 16p11.2 microdeletion, 10 carried an additional large copy-number variant, a significantly higher proportion compared with controls conditioned for a large first hit (10 of 42 cases, 21 of 471 controls; $p = 0.000057$, odds ratio $= 6.6$) (49). The clinical features of individuals with two mutations were distinct from and/or more severe than those of individuals carrying only the co-occurring mutation. Another study showed that three individuals with ASD who carried a de novo *SHANK2* deletion were also carriers of a second hit at the 15q11 locus (74). Two were carrying copy-number variants between BP4 and BP5, including *CHRNA7* and *ARHGAP11B*; the third was carrying a deletion between BP1 and BP2 that removed the *CYFIP1*, *NIPA1*, *NIPA2*, and *TUBGCP5* genes. Finally, Sato et al. (102) identified a *SHANK1* deletion on chromosome 19 that cosegregated with high-functioning autism in males but not in females. A stop mutation of the protocadherin gamma 11–encoding gene (*PCDHGA11*) was identified in all individuals, suggesting that, similarly to individuals with *SHANK2* mutations, additional hits increase the risk of ASD in individuals with *SHANK1* mutations. Because of the small sample, it is not possible to confirm whether genetic interaction between *SHANK1*, *SHANK2*, and other synaptic genes is necessary to develop ASD. Nevertheless, these results should prompt researchers to identify putative modifier genes even in individuals carrying de novo mutations.

As in other complex genetic disorders, the causative role of the inherited variants is difficult to ascertain. Interestingly, combined genetic and functional studies of several synaptic genes have repeatedly shown an enrichment of deleterious inherited variants in individuals with ASD. For example, inherited variants of *NLGN3*, *NLGN4X*, *SHANK2*, *SHANK3*, and *NRXN1* decrease synaptic density in vitro (6, 15, 21, 25, 36, 37, 74). Why these variants have a phenotypic impact in the affected individuals and not in the parents is unknown. Several recessive cases of ASD have been reported, but mostly in individuals with metabolic disorders (reviewed in 27) or in consanguineous families (83, 135). In a cohort of 933 cases and 869 controls, Lim et al. (76) recently identified a twofold increase of deleterious mutations affecting the two copies of autosomal genes that normally display low rates of loss-of-function variation ($\leq$5% frequency). On the X chromosome, a similar 1.5-fold increase in rare hemizygous knockouts was observed in males with ASD. Taken together, 3% of autosomal recessive mutations contribute to ASD and another 2% of X-linked mutations could be associated with ASD in males.

To identify candidate genomic regions containing recessive variations, Casey et al. (22) used homozygous haplotype mapping of 1,402 trios genotyped for 1 million SNPs. They determined that regions of homozygous haplotypes were significantly enriched in regions previously reported for ASD, suggesting the presence of unidentified recessive mutations. Another study

Figure 5

Biological pathways of proteins mutated in ASD or other disorders. These proteins are involved in many functions, including the organization of the postsynaptic density, cytoskeleton dynamics, cellular signaling cascades, epigenetic regulation of transcription, and release of neurotransmitters. Proteins associated with ASD are in red, those associated with other psychiatric disorders are in purple, and those associated with intellectual disability are in blue. Panel *a* illustrates the pre- and postsynapse; panel *b* illustrates the synapse and gene transcription regulation in the nucleus.



identified both a copy-number variant and a coding-sequence variant in an individual with a maternally inherited *DIAPH3* deletion and a paternally inherited *DIAPH3* P614T deleterious mutation affecting the number of cellular filopodia in vitro (124). Studies are currently under way to determine whether copy-number variants could unmask mutations of the remaining allele in other individuals with ASD.

These results indicate that recessive cases of ASD exist and that others will certainly be reported in the future. However, it remains striking that the analysis of the copy-number variants could not detect a larger number of individuals carrying recessive mutations (e.g., two deletions of the same gene), suggesting that this mode of inheritance is not preponderant in ASD. Mutations affecting neuronal genes on both alleles might be associated with more severe syndromes than ASD, as observed in homozygous mutations of *CNTNAP2* or *NRXN1* that cause an autosomal recessive Pitt-Hopkins-like syndrome characterized by severe intellectual deficiency, epilepsy, and breathing anomalies (136).

In summary, the presence of multiple hits at different loci might be the rule rather than the exception in ASD. One challenge is to identify the combination of susceptibility genes that either additively or in interaction (epistasis) cause the disease. Identification of protective alleles could also shed light on new targets for treatments.

COMMON VARIANTS IN ASD

Owing to the absence of classical Mendelian inheritance, autism was first thought to be a polygenic trait involving many common variants, each with a small effect. Model-free linkage studies, such as affected sib-pair analyses, were therefore performed and followed by genome-wide association studies (GWAS). However, these approaches identified only a limited number of loci, and the results were rarely replicated (4, 114, 127, 128). The lack of significant results obtained through GWAS—which test each SNP individually—has thus raised the question of the role of common variants in ASD susceptibility. To better understand the missing heritability of complex traits in humans (79), Visscher's group (133) introduced a new method that considers all genotyped SNPs simultaneously to estimate the proportion of variance for a trait explained by these common variants. For height and IQ, this method revealed that a large proportion of heritability was captured by SNPs, even though only few significant single-SNP signals have been obtained by GWAS (30, 133). Klei et al. (67) recently used this quantitative genetics approach to estimate the heritability of ASD in a large sample of more than 2,000 families with ASD and 3,600 controls. They found substantial additive genetic effects of SNPs on ASD liability, explaining more than 60% of the variance for individuals with ASD originating from multiplex families and approximately 40% for those from simplex families. The advantage of this method is that it shows that SNPs can actually capture a relatively large fraction of ASD heritability; the disadvantage is that it does not provide information on the genes involved.

Alternative methodologies have been used to aggregate the variants of interest (e.g., genes in a given biological pathway) and generate a predictive score for ASD diagnosis (109, 134). The first results obtained by such methodologies are encouraging because they correctly predicted ASD diagnosis in 56–85% of the cases (109). However, as for any association studies, these initial positive results need to be confirmed in large cohorts of individuals before the "predictive scores" are used in diagnostic tests.

In summary, common variants such as SNPs play a role in ASD susceptibility, but individually they are associated with very small effects and are widely distributed across the genome. The relevance of the "predictive scores" based on common SNPs for current diagnosis must be confirmed

in large samples. In the future, however, a better ascertainment of the genetic background of each individual might be of great importance in better predicting clinical outcomes.

GENETIC PATHWAYS ASSOCIATED WITH ASD

Several databases provide functional annotation of genes associated with ASD. For this review, we used two databases that are updated regularly: AutismKB (**http://autismkb.cbi.pku.edu.cn**) and SFARI Gene (**https://gene.sfari.org**). A total of 197 genes are included in both databases, and 481 are additionally included in either one or the other (255 in AutismKB and 226 in SFARI Gene). The main difference between the two databases concerns the selection of the genes. AutismKB usually selects genes from linkage analyses, copy-number variant studies, and GWAS, whereas SFARI Gene usually selects genes from copy-number variant studies, sequencing analyses of large cohorts, and case reports. **Figure 6** illustrates the findings for 203 genes with de novo mutations identified in individuals with ASD. A preliminary functional annotation clustering analysis using GeneMANIA (**http://www.genemania.org**) and DAVID (**http://david.abcc.ncifcrf.gov**) indicates that 36% of the proteins have at least one interaction (direct or indirect) with another protein, 61% are expressed in the brain, and 14% are known to be involved in synaptic function.

Monogenic forms of ASD include genetic diseases such as tuberous sclerosis (associated with *TSC1* and *TSC2*), neurofibromatosis (associated with *NF1*), Rett syndrome (associated with *MECP2*), and Cowden syndrome (associated with *PTEN*). *FMR1*, the gene mutated in fragile X syndrome, is the single gene whose mutation is most commonly observed in ASD, accounting for ~2% of cases. Rare mutations have also been identified in synaptic genes, including *NLGN3* and *NLGN4X* (62); *SHANK1*, *SHANK2*, and *SHANK3* (36, 74, 102); *NRXN1* (114); and *CNTNAP2* (5, 10). More recently, studies of mutations in *CHD8*, which is associated with chromatin remodeling (89, 90), and *TMLHE*, which catalyzes carnitine biosynthesis (23, 85), confirmed that the biological functions associated with ASD go beyond those involved directly in synaptic function and affect a wide range of cellular processes.

Several pathway analyses have been performed using either genetic or transcriptome data to gain insight into the biological functions associated with ASD. Pinto et al. (96) analyzed the burden of copy-number variants in 996 individuals with ASD and found an enrichment of genomic rearrangements disrupting genes involved in cellular proliferation, cellular projection and motility, and GTPase/Ras signaling. Iossifov et al. (57) analyzed the exome sequencing data of 343 families with ASD and observed that de novo mutations were enriched in genes coding for proteins associated with the fragile X FMRP protein pathway. Voineagu et al. (122) analyzed genes that are differentially expressed between two brain regions (frontal and temporal lobes) in individuals with ASD and controls. Interestingly, the typical regional differences between the gene expression profiles of the frontal and temporal lobes were attenuated in individuals with ASD. A first network module was related to interneurons and to genes involved in synaptic function, and was downregulated in brains from individuals with ASD compared with those from controls; a second module was enriched for genes related to immunity and microglia activation, and was upregulated in brains from individuals with ASD compared with those from controls.

Based on these genetic results and studies of animal models (see below), two pathways are emerging (18). The first pathway was suggested by reports of aberrant mRNA translation of synaptic proteins (64). The evidence supporting this theory includes the identification of mutations affecting several proteins (TSC1, TSC2, NF1, and PTEN) that normally inhibit translation through the PI3K-mTOR signaling pathway as well as mutations affecting proteins directly involved in inhibiting mRNA translation at the synapse (FMRP, CYFIP1, and EIF4E) (**Figure 5**). The second pathway concerns the excitation/inhibition balance. Several genes associated with

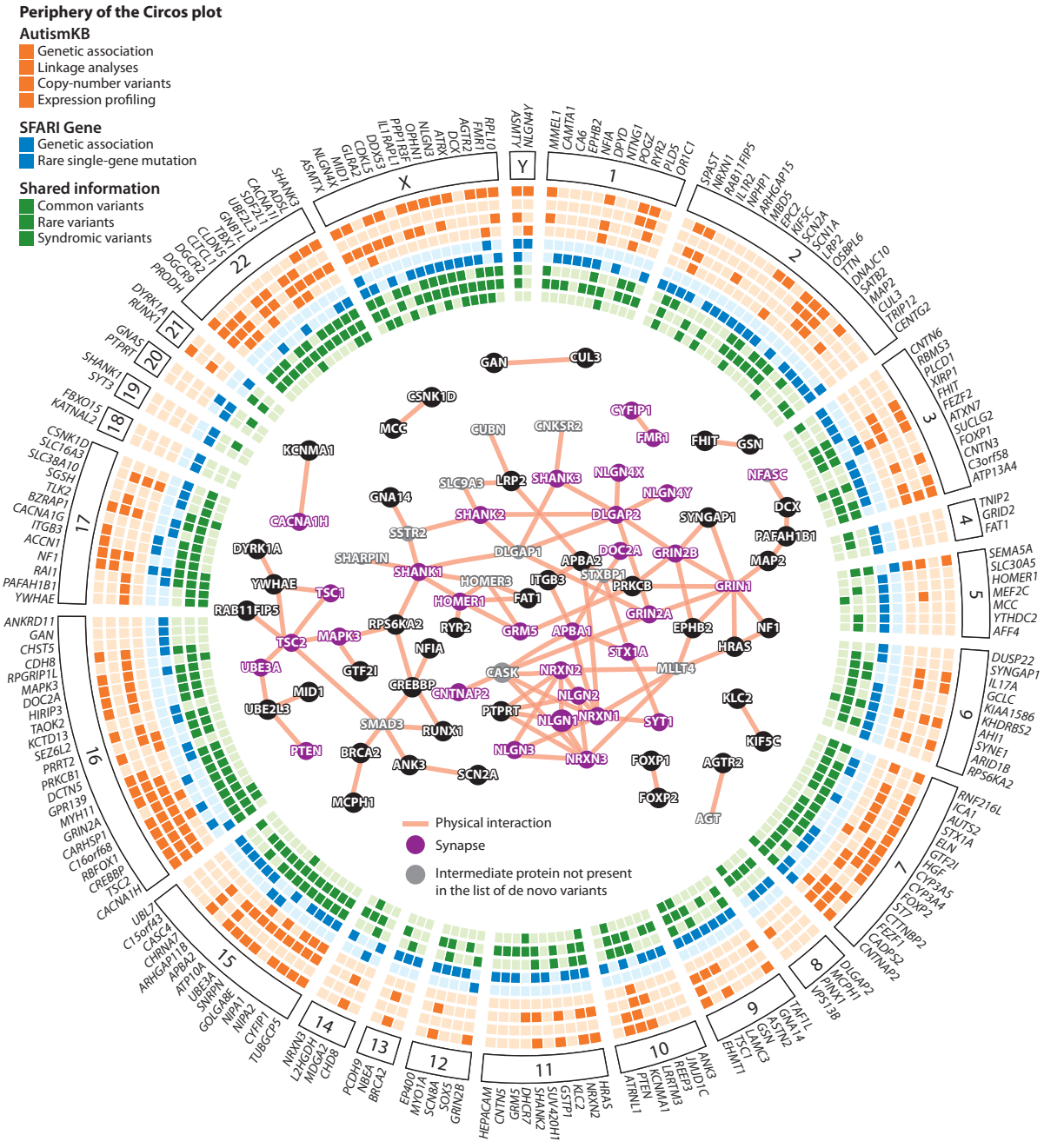


Figure 6

Circos plot of de novo mutations in ASD. All coding-sequence variants and copy-number variants present in AutismKB (updated July 2011) and SFARI Gene (updated September 2012) are shown. A GeneMANIA network analysis (*center*) highlights proteins with synaptic function.

ASD, such as *NLGN*, *NRXN*, and *SHANK*, appear to be involved in the formation of excitatory and inhibitory synapses. In addition, genes associated with epilepsy, such as *SCN1A*, which encodes a voltage-gated sodium channel, were also found to be mutated in individuals with ASD. These two pathways—mRNA translation and the excitation/inhibition balance—are potential drug targets, and clinical trials are ongoing to determine whether targeting them could improve the symptoms of individuals with ASD (60).

In summary, the genes associated with ASD are numerous and involved in multiple cellular functions, including chromatin remodeling, metabolism, mRNA translation, and synaptic function. The downstream consequences of the mutations, however, might converge to a defect in neuronal/synaptic homeostasis (97, 116) (**Figure 5**).

MOUSE MODELS OF ASD

A large number of mouse models have been developed that have construct validity for the pathways highlighted by genetic studies: PTEN, UBE3A, FMR1, MECP2, mTOR, TSC, and cell adhesion molecules and scaffolding proteins localized specifically at the synapse (40) (**Figure 7**). The behavioral characterization of these models supports the involvement of these genes in the core symptoms of ASD (40). The mice usually display abnormalities in social interactions (caged and free interactions, same-sex and cross-sex interactions), communication (ultrasonic vocalizations, scent marking), and repetitive behaviors (self-grooming, jumping, marble burying) (108). Interestingly, models carrying mutations in *Mecp2*, *Cntnap2*, *Shank2*, and *Shank3* display abnormalities in all three domains, suggesting that these genes play a major role in circuits involved in both social communication and repetitive behaviors (93, 94, 104, 131). *Fmr1*, *Shank2*, and *Cntnap2* knockout mice also display hyperactivity, a common feature in individuals with ASD (38, 104, 131).

At the cellular or network level, both decreases and increases of synaptic density have been observed. For example, a reduction of dendrites was observed in Rett syndrome and tuberous sclerosis, in contrast to the high density of dendrites reported in fragile X syndrome (8). Similarly, some mutations appeared to increase glutamate currents, whereas others reduced them (104), and the same has been observed for inhibitory currents (115). Interestingly, several genes associated with ASD are modulated by neuronal activity, suggesting that they may mediate experience-dependent circuit modifications (116).

Recent studies have also shown dramatic improvements in synaptic function and behavior even in adult mice, suggesting that, at least in some cases, the causative gene does not play a pivotal role in brain development but may instead be required to maintain full neurological function once development is complete (13, 52). One pioneer experiment was performed by Bird's group (52), who created a mouse in which the endogenous *Mecp2* gene was silenced but could be conditionally activated. Using this tool, they demonstrated that activation of *Mecp2* in ~70% of neurons in both immature and mature adult animals reversed many morphological defects in the motor cortex, including defects in neuronal size and dendritic complexity, and led to significant improvements in respiratory and sensory-motor functions, including breathing patterns, grip strength, and balance beam and rotarod performance (52). More recently, Derecki et al. (33) used transplantation of wild-type bone marrow to introduce wild-type microglial cells in the Rett mouse model and were able to arrest disease development: Life span was increased, breathing patterns were normalized, apneas were reduced, body weight was increased to near that of wild type, and locomotor activity was improved. Finally, *Nlgn3* knockout mice exhibit ectopic synapse formation and perturbed mGluR-dependent synaptic plasticity, a hallmark of fragile X syndrome. These phenotypes were rescued by reexpression of *Nlgn3* in juvenile mice, again highlighting the possibility of reverting neuronal circuit alterations in autism after the completion of development (13).

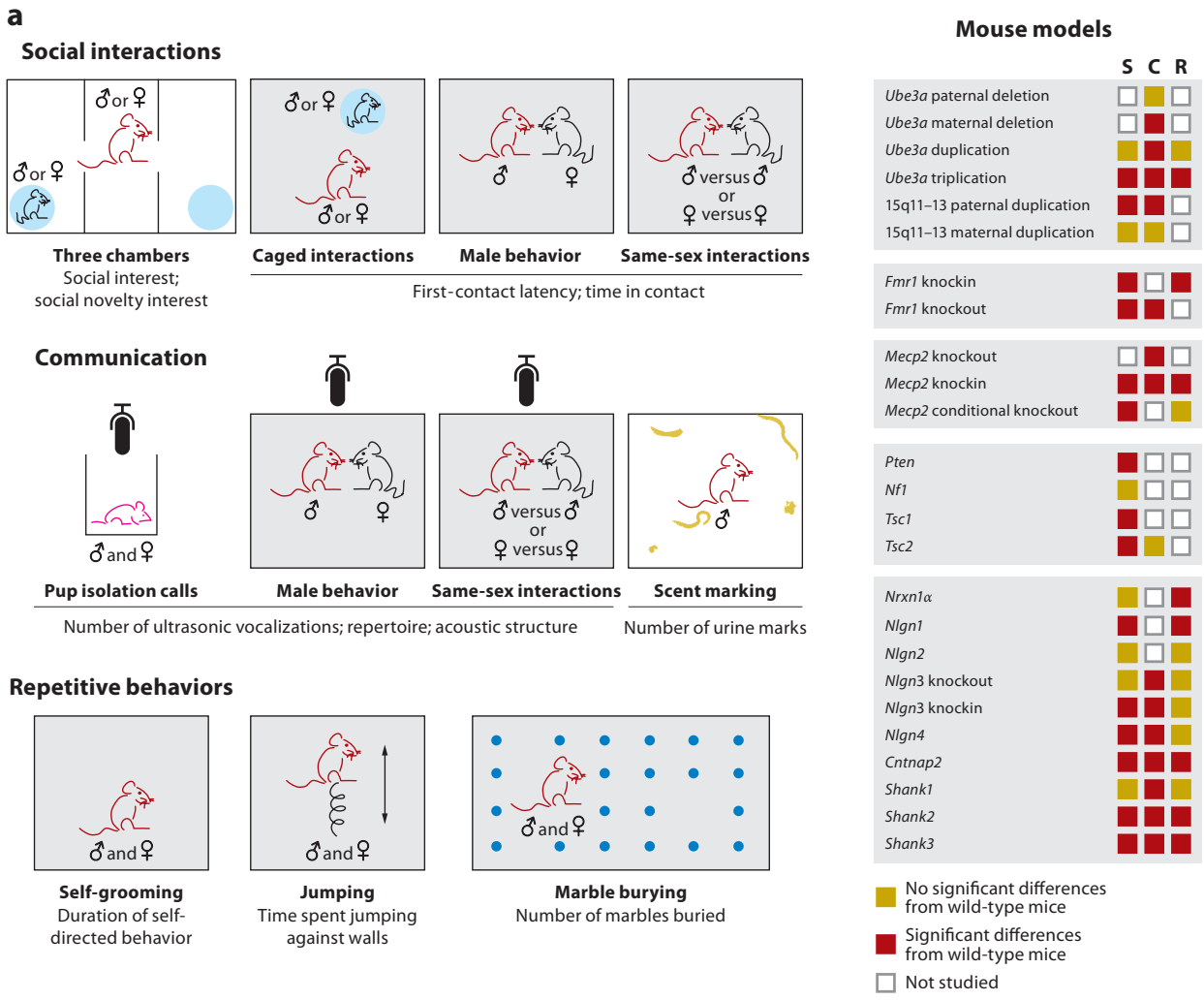


Figure 7

Behavioral characterization of mouse models of ASD: methods and results. (*a*) The main behavioral tests used to evaluate the equivalents of the three core ASD symptoms in mouse models: social interactions (caged and free interactions, same-sex and cross-sex interactions), communication (ultrasonic vocalizations, scent marking), and repetitive behaviors (self-grooming, jumping, marble burying). (*b*) Recapitulation of presence or absence of deficits in the equivalents of the three core ASD symptoms (S, social interactions; C, communication; R, repetitive behaviors) in mouse models of ASD.

Finally, even in genetically homogeneous mouse strains, phenotypic variability can still be observed. For instance, *Nlgn4* knockout and *Nlgn3*-R451C mice that initially displayed deficits in social interactions and/or ultrasonic communication (63, 114) did not show such deficits in later studies (24, 41). This variability is not surprising because individuals with ASD who carry identical mutations also display different clinical outcomes [e.g., individuals with *NLGN4X* mutations may or may not display intellectual disability (62, 72)]. The next step to better understand this variability is to study the role of epigenetic and environmental factors and to create new mouse models carrying two or more mutations in different genes. These models would be closer to the clinical reality of individuals carrying multiple mutations.

In summary, mouse models carrying mutations in pathways associated with ASD can display abnormalities in social interactions, communication, and repetitive behaviors. The phenotypic variability highlighted in the different models is reminiscent of the clinical heterogeneity observed in individuals with ASD. The idea that ASD phenotypes are stable and insensitive to treatment is now challenged. Cellular and animal models, together with knowledge-based synaptic clinical trials, should hopefully tell us which affected genes/pathways can be efficiently recovered.

PROSPECTS FOR THE GENETICS OF ASD

In the past decade, important discoveries have been made in the genetics of ASD with the identification of a large number of causative genes that converge in common pathways, such as synaptic mRNA translation and dendrite/synapse formation. However, many questions remain unanswered. For example, the role of the inherited variants remains difficult to ascertain even though they certainly contribute to a large fraction of genetic ASD susceptibility. The excess of males in ASD remains largely unexplained; females seem to buffer deleterious copy-number variants and coding-sequence variants more effectively than males (who, in a way by definition, are carriers of a first hit: the complete absence of one X chromosome).

But, besides genetics, one of the main challenges in the field of ASD concerns the phenotypes associated with each causative gene or combination of susceptibility genes. The clinical outcomes associated with the causative genes exceed the boundaries of ASD because the same genes associated with ASD (e.g., *NRXN1*, *SHANK3*, and *CNTNAP2*) are also associated with other neuropsychiatric disorders, such as schizophrenia and bipolar disorder (45, 65, 99, 121, 125). This pleiotropy raises the questions of the role each gene plays in the development and function of the human brain and the extent to which different combinations of susceptibility alleles could predict clinical outcomes. Collaborative work involving clinicians, geneticists, and neurobiologists is therefore necessary to better understand the diversity of individuals with ASD and to improve their diagnosis, care, and integration.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We thank Roberto Toro for critical reading of the manuscript and Alexandre Mathieu for the figures. This work was funded by the Institut Pasteur, the Bettencourt-Schueller foundation, Centre National de la Recherche Scientifique, Université Paris Diderot, Agence Nationale de la Recherche (ANR-08-MNPS-037-01-SynGen), Neuron-ERANET (EUHF-AUTISM), the Conny-Maeva Charitable Foundation, the Cognacq-Jay Foundation, the Orange Foundation, the Fondamental Foundation, and the Innovative Medicines Initiative Joint Undertaking under grant agreement 115300, resources of which are composed of financial contributions from the European Union's Seventh Framework Programme (FP7/2007–2013) and European Federation of Pharmaceutical Industries and Associations (EFPIA) companies' in-kind contributions.

LITERATURE CITED

1. Abrahams BS, Geschwind DH. 2008. Advances in autism genetics: on the threshold of a new neurobiology. *Nat. Rev. Genet.* 9:341–55

2. Alkan C, Coe BP, Eichler EE. 2011. Genome structural variation discovery and genotyping. *Nat. Rev. Genet.* 12:363–76
3. Amir RE, Van den Veyver IB, Wan M, Tran CQ, Francke U, Zoghbi HY. 1999. Rett syndrome is caused by mutations in X-linked *MECP2*, encoding methyl-CpG-binding protein 2. *Nat. Genet.* 23:185–88
4. Anney R, Klei L, Pinto D, Almeida J, Bacchelli E, et al. 2012. Individual common variants exert weak effects on the risk for autism spectrum disorders. *Hum. Mol. Genet.* 21:4781–92
5. Arking DE, Cutler DJ, Brune CW, Teslovich TM, West K, et al. 2008. A common genetic variant in the neurexin superfamily member *CNTNAP2* increases familial risk of autism. *Am. J. Hum. Genet.* 82:160–64
6. Arons MH, Thynne CJ, Grabrucker AM, Li D, Schoen M, et al. 2012. Autism-associated mutations in ProSAP2/Shank3 impair synaptic transmission and neurexin-neuroligin-mediated transsynaptic signaling. *J. Neurosci.* 32:14966–78
7. Asperger H. 1944. Die "autistischen Psychopathen" im Kindesalter. *Arch. Psychiatr. Nervenkrankh.* 177:76–137
8. Auerbach BD, Osterweil EK, Bear MF. 2011. Mutations causing syndromic autism define an axis of synaptic pathophysiology. *Nature* 480:63–68
9. Bailey A, Le Couteur A, Gottesman I, Bolton P, Simonoff E, et al. 1995. Autism as a strongly genetic disorder: evidence from a British twin study. *Psychol. Med.* 25:63–77
10. Bakkaloglu B, O'Roak BJ, Louvi A, Gupta AR, Abelson JF, et al. 2008. Molecular cytogenetic analysis and resequencing of contactin associated protein-like 2 in autism spectrum disorders. *Am. J. Hum. Genet.* 82:165–73
11. Ballif BC, Hornor SA, Jenkins E, Madan-Khetarpal S, Surti U, et al. 2007. Discovery of a previously unrecognized microdeletion syndrome of 16p11.2–p12.2. *Nat. Genet.* 39:1071–73
12. Battaglia A, Novelli A, Bernardini L, Igliozzi R, Parrini B. 2009. Further characterization of the new microdeletion syndrome of 16p11.2–p12.2. *Am. J. Med. Genet. A* 149A:1200–4
13. Baudouin SJ, Gaudias J, Gerharz S, Hatstatt L, Zhou K, et al. 2012. Shared synaptic pathophysiology in syndromic and nonsyndromic rodent models of autism. *Science* 338:128–32
14. Ben-Shachar S, Lanpher B, German JR, Qasaymeh M, Potocki L, et al. 2009. Microdeletion 15q13.3: a locus with incomplete penetrance for autism, mental retardation, and psychiatric disorders. *J. Med. Genet.* 46:382–88
15. Berkel S, Tang W, Trevino M, Vogt M, Obenhaus HA, et al. 2011. Inherited and de novo SHANK2 variants associated with autism spectrum disorder impair neuronal morphogenesis and physiology. *Hum. Mol. Genet.* 21:344–57
16. Bijlsma EK, Gijsbers AC, Schuurs-Hoeijmakers JH, van Haeringen A, Fransen van de Putte DE, et al. 2009. Extending the phenotype of recurrent rearrangements of 16p11.2: deletions in mentally retarded patients without autism and in normal individuals. *Eur. J. Med. Genet.* 52:77–87
17. Bittel DC, Kibiryeva N, Butler MG. 2006. Expression of 4 genes between chromosome 15 breakpoints 1 and 2 and behavioral outcomes in Prader-Willi syndrome. *Pediatrics* 118:e1276–83
18. Bourgeron T. 2009. A synaptic trek to autism. *Curr. Opin. Neurobiol.* 19:231–34
19. Burnside RD, Pasion R, Mikhail FM, Carroll AJ, Robin NH, et al. 2011. Microdeletion/microduplication of proximal 15q11.2 between BP1 and BP2: a susceptibility region for neurological dysfunction including developmental and language delay. *Hum. Genet.* 130:517–28
20. Butler MG, Fischer W, Kibiryeva N, Bittel DC. 2008. Array comparative genomic hybridization (aCGH) analysis in Prader-Willi syndrome. *Am. J. Med. Genet. A* 146A:854–60
21. Camacho-Garcia RJ, Planelles MI, Margalef M, Pecero ML, Martínez-Leal R, et al. 2012. Mutations affecting synaptic levels of neurexin-1 β in autism and mental retardation. *Neurobiol. Dis.* 47:135–43
22. Casey JP, Magalhaes T, Conroy JM, Regan R, Shah N, et al. 2011. A novel approach of homozygous haplotype sharing identifies candidate genes in autism spectrum disorder. *Hum. Genet.* 131:565–79
23. Celestino-Soper PB, Violante S, Crawford EL, Luo R, Lionel AC, et al. 2012. A common X-linked inborn error of carnitine biosynthesis may be a risk factor for nondysmorphic autism. *Proc. Natl. Acad. Sci. USA* 109:7974–81
24. Chadman KK, Gong S, Scattoni ML, Boltuck SE, Gandhy SU, et al. 2008. Minimal aberrant behavioural phenotypes of neuroligin-3 R451C knockin mice. *Autism Res.* 1:147–58

25. Chih B, Afridi SK, Clark L, Scheiffele P. 2004. Disorder-associated mutations lead to functional inactivation of neuroligins. *Hum. Mol. Genet.* 13:1471–77
26. Christian SL, Brune CW, Sudi J, Kumar RA, Liu S, et al. 2008. Novel submicroscopic chromosomal abnormalities detected in autism spectrum disorder. *Biol. Psychiatry* 63:1111–17
27. Coleman M, Gillberg C. 2012. *The Autisms*. Oxford, UK: Oxford Univ. Press. 432 pp.
28. Cook EH Jr, Lindgren V, Leventhal BL, Courchesne R, Lincoln A, et al. 1997. Autism or atypical autism in maternally but not paternally derived proximal 15q duplication. *Am. J. Hum. Genet.* 60:928–34
29. Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, et al. 2011. A copy number variation morbidity map of developmental delay. *Nat. Genet.* 43:838–46
30. Davies G, Tenesa A, Payton A, Yang J, Harris SE, et al. 2011. Genome-wide association studies establish that human intelligence is highly heritable and polygenic. *Mol. Psychiatry* 16:996–1005
31. Depienne C, Héron D, Betancur C, Benyahia B, Trouillard O, et al. 2007. Autism, language delay and mental retardation in a patient with 7q11 duplication. *J. Med. Genet.* 44:452–58
32. Depienne C, Moreno-De-Luca D, Héron D, Bouteiller D, Gennetier A, et al. 2009. Screening for genomic rearrangements and methylation abnormalities of the 15q11-q13 region in autism spectrum disorders. *Biol. Psychiatry* 66:349–59
33. Derecki NC, Cronk JC, Lu Z, Xu E, Abbott SB, et al. 2012. Wild-type microglia arrest pathology in a mouse model of Rett syndrome. *Nature* 484:105–9
34. Devlin B, Scherer SW. 2012. Genetic architecture in autism spectrum disorder. *Curr. Opin. Genet. Dev.* 22:229–37
35. Dibbens LM, Mullen S, Helbig I, Mefford HC, Bayly MA, et al. 2009. Familial and sporadic 15q13.3 microdeletions in idiopathic generalized epilepsy: precedent for disorders with complex inheritance. *Hum. Mol. Genet.* 18:3626–31
36. Durand CM, Betancur C, Boeckers TM, Bockmann J, Chaste P, et al. 2007. Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. *Nat. Genet.* 39:25–27
37. Durand CM, Perroy J, Loll F, Perrais D, Fagni L, et al. 2011. *SHANK3* mutations identified in autism lead to modification of dendritic spine morphology via an actin-dependent mechanism. *Mol. Psychiatry* 17:71–84
38. Dutch-Belg. Fragile X Consort. 1994. *Fmr1* knockout mice: a model to study fragile X mental retardation. *Cell* 78:23–33
39. Elsabbagh M, Divan G, Koh YJ, Kim YS, Kauchali S, et al. 2012. Global prevalence of autism and other pervasive developmental disorders. *Autism Res.* 5:160–79
40. Ey E, Leblond CS, Bourgeron T. 2010. Behavioral profiles of mouse models for autism spectrum disorders. *Autism Res.* 4:5–16
41. Ey E, Yang M, Katz AM, Woldeyohannes L, Silverman JL, et al. 2012. Absence of deficits in social behaviors and ultrasonic vocalizations in later generations of mice lacking neuroligin4. *Genes Brain Behav.* 11:928–41
42. Fernandez BA, Roberts W, Chung B, Weksberg R, Meyn S, et al. 2010. Phenotypic spectrum associated with de novo and inherited deletions and duplications at 16p11.2 in individuals ascertained for diagnosis of autism spectrum disorder. *J. Med. Genet.* 47:195–203
43. Folstein S, Rutter M. 1977. Infantile autism: a genetic study of 21 twin pairs. *J. Child Psychol. Psychiatry* 18:297–321
44. Freitag CM. 2007. The genetics of autistic disorders and its clinical relevance: a review of the literature. *Mol. Psychiatry* 12:2–22
45. Gauthier J, Champagne N, Lafrenière RG, Xiong L, Spiegelman D, et al. 2010. De novo mutations in the gene encoding the synaptic scaffolding protein *SHANK3* in patients ascertained for schizophrenia. *Proc. Natl. Acad. Sci. USA* 107:7863–68
46. Gillberg C. 2010. The ESSENCE in child psychiatry: Early Symptomatic Syndromes Eliciting Neurodevelopmental Clinical Examinations. *Res. Dev. Disabil.* 31:1543–51
47. Gilman SR, Iossifov I, Levy D, Ronemus M, Wigler M, Vitkup D. 2011. Rare de novo variants associated with autism implicate a large functional network of genes involved in formation and function of synapses. *Neuron* 70:898–907

48. Girirajan S, Rosenfeld JA, Coe BP, Parikh S, Friedman N, et al. 2012. Phenotypic heterogeneity of genomic disorders and rare copy-number variants. *N. Engl. J. Med.* 367:1321–31
49. Girirajan S, Rosenfeld JA, Cooper GM, Antonacci F, Siswara P, et al. 2010. A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay. *Nat. Genet.* 42:203–9
50. Glessner JT, Wang K, Cai G, Korvatska O, Kim CE, et al. 2009. Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature* 459:569–73
51. Golzio C, Willer J, Talkowski ME, Oh EC, Taniguchi Y, et al. 2012. *KCTD13* is a major driver of mirrored neuroanatomical phenotypes of the 16p11.2 copy number variant. *Nature* 485:363–67
52. Guy J, Gan J, Selfridge J, Cobb S, Bird A. 2007. Reversal of neurological defects in a mouse model of Rett syndrome. *Science* 315:1143–47
53. Hallmayer J, Cleveland S, Torres A, Phillips J, Cohen B, et al. 2011. Genetic heritability and shared environmental factors among twin pairs with autism. *Arch. Gen. Psychiatry* 68:1095–102
54. Hartley SL, Maclean WE Jr, Butler MG, Zarcone J, Thompson T. 2005. Maladaptive behaviors and risk factors among the genetic subtypes of Prader-Willi syndrome. *Am. J. Med. Genet. A* 136A:140–45
55. Helbig I, Mefford HC, Sharp AJ, Guipponi M, Fichera M, et al. 2009. 15q13.3 microdeletions increase risk of idiopathic generalized epilepsy. *Nat. Genet.* 41:160–62
56. Int. Schizophr. Consort. 2008. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* 455:237–41
57. Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, et al. 2012. De novo gene disruptions in children on the autistic spectrum. *Neuron* 74:285–99
58. Itsara A, Wu H, Smith JD, Nickerson DA, Romieu I, et al. 2010. De novo rates and selection of large copy number variation. *Genome Res.* 20:1469–81
59. Jacquemont ML, Sanlaville D, Redon R, Raoul O, Cormier-Daire V, et al. 2006. Array-based comparative genomic hybridisation identifies high frequency of cryptic chromosomal rearrangements in patients with syndromic autism spectrum disorders. *J. Med. Genet.* 43:843–49
60. Jacquemont S, Curie A, des Portes V, Torrioli MG, Berry-Kravis E, et al. 2011. Epigenetic modification of the *FMR1* gene in fragile X syndrome is associated with differential response to the mGluR5 antagonist AFQ056. *Sci. Transl. Med.* 3:64ra1
61. Jacquemont S, Reymond A, Zufferey F, Harewood L, Walters RG, et al. 2011. Mirror extreme BMI phenotypes associated with gene dosage at the chromosome 16p11.2 locus. *Nature* 478:97–102
62. Jamain S, Quach H, Betancur C, Råstam M, Colineaux C, et al. 2003. Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. *Nat. Genet.* 34:27–29
63. Kanner L. 1943. Autistic disturbances of affective contact. *Nerv. Child* 2:217–50
64. Kelleher RJ, Bear MF. 2008. The autistic neuron: troubled translation? *Cell* 135:401–6
65. Kirov G, Gumus D, Chen W, Norton N, Georgieva L, et al. 2008. Comparative genome hybridization suggests a role for *NRXN1* and *APBA2* in schizophrenia. *Hum. Mol. Genet.* 17:458–65
66. Kishino T, Lalande M, Wagstaff J. 1997. UBE3A/E6-AP mutations cause Angelman syndrome. *Nat. Genet.* 15:70–73
67. Klei L, Sanders SJ, Murtha MT, Hus V, Lowe JK, et al. 2012. Common genetic variants, acting additively, are a major source of risk for autism. *Mol. Autism* 3:9
68. Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, et al. 2012. Rate of de novo mutations and the importance of father's age to disease risk. *Nature* 488:471–75
69. Konyukh M, Delorme R, Chaste P, Leblond C, Lemière N, et al. 2011. Variations of the candidate *SEZ6L2* gene on chromosome 16p11.2 in patients with autism spectrum disorders and in human populations. *PLoS ONE* 6:e17289
70. Kumar RA, KaraMohamed S, Sudi J, Conrad DF, Brune C, et al. 2008. Recurrent 16p11.2 microdeletions in autism. *Hum. Mol. Genet.* 17:628–38
71. Kumar RA, Marshall CR, Badner JA, Babatz TD, Mukamel Z, et al. 2009. Association and mutation analyses of 16p11.2 autism candidate genes. *PLoS ONE* 4:e4582
72. Laumonnier F, Bonnet-Brilhault F, Gomot M, Blanc R, David A, et al. 2004. X-linked mental retardation and autism are associated with a mutation in the *NLGN4* gene, a member of the neuroligin family. *Am. J. Hum. Genet.* 74:552–57

73. LeBlanc JJ, Fagiolini M. 2011. Autism: a "critical period" disorder? *Neural Plast.* 2011:921680
74. Leblond CS, Heinrich J, Delorme R, Proepper C, Betancur C, et al. 2012. Genetic and functional analyses of *SHANK2* mutations suggest a multiple hit model of autism spectrum disorders. *PLoS Genet.* 8:e1002521
75. Lichtenstein P, Carlstrom E, Råstam M, Gillberg C, Anckarsater H. 2010. The genetics of autism spectrum disorders and related neuropsychiatric disorders in childhood. *Am. J. Psychiatry* 167:1357–63
76. Lim ET, Raychaudhuri S, Sanders SJ, Stevens C, Sabo A, et al. 2013. Rare complete knockouts in humans: population distribution and significant role in autism spectrum disorders. *Neuron* 77:235–42
77. Lundstrom S, Chang Z, Råstam M, Gillberg C, Larsson H, et al. 2012. Autism spectrum disorders and autistic like traits: similar etiology in the extreme end and the normal variation. *Arch. Gen. Psychiatry* 69:46–52
78. Mabb AM, Judson MC, Zylka MJ, Philpot BD. 2011. Angelman syndrome: insights into genomic imprinting and neurodevelopmental phenotypes. *Trends Neurosci.* 34:293–303
79. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, et al. 2009. Finding the missing heritability of complex diseases. *Nature* 461:747–53
80. Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, et al. 2008. Structural variation of chromosomes in autism spectrum disorder. *Am. J. Hum. Genet.* 82:477–88
81. Matsuura T, Sutcliffe JS, Fang P, Galjaard RJ, Jiang YH, et al. 1997. De novo truncating mutations in E6-AP ubiquitin-protein ligase gene (*UBE3A*) in Angelman syndrome. *Nat. Genet.* 15:74–77
82. Miller DT, Shen Y, Weiss LA, Korn J, Anselm I, et al. 2009. Microdeletion/duplication at 15q13.2q13.3 among individuals with features of autism and other neuropsychiatric disorders. *J. Med. Genet.* 46:242–48
83. Morrow EM, Yoo SY, Flavell SW, Kim TK, Lin Y, et al. 2008. Identifying autism loci and genes by tracing recent shared ancestry. *Science* 321:218–23
84. Napoli I, Mercaldo V, Boyl PP, Eleuteri B, Zalfa F, et al. 2008. The fragile X syndrome protein represses activity-dependent translation through CYFIP1, a new 4E-BP. *Cell* 134:1042–54
85. Nava C, Lamari F, Héron D, Mignot C, Rastetter A, et al. 2012. Analysis of the chromosome X exome in patients with autism spectrum disorders identified novel candidate genes, including *TMLHE*. *Transl. Psychiatry* 2:e179
86. Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, et al. 2012. Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* 485:242–45
87. Neves-Pereira M, Müller B, Massie D, Williams JH, O'Brien PC, et al. 2009. Deregulation of *EIF4E*: a novel mechanism for autism. *J. Med. Genet.* 46:759–65
88. O'Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, et al. 2011. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nat. Genet.* 43:585–89
89. O'Roak BJ, Vives L, Fu W, Egertson JD, Stanaway IB, et al. 2012. Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science* 338:1619–22
90. O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, et al. 2012. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* 485:246–50
91. Ozonoff S, Young GS, Carter A, Messinger D, Yirmiya N, et al. 2011. Recurrence risk for autism spectrum disorders: a Baby Siblings Research Consortium study. *Pediatrics* 128:e488–95
92. Pagnamenta AT, Wing K, Akha ES, Knight SJ, Bolte S, et al. 2009. A 15q13.3 microdeletion segregating with autism. *Eur. J. Hum. Genet.* 17:687–92
93. Peça J, Feliciano C, Ting JT, Wang W, Wells MF, et al. 2011. *Shank3* mutant mice display autistic-like behaviours and striatal dysfunction. *Nature* 472:437–42
94. Peñagarikano O, Abrahams BS, Herman EI, Winden KD, Gdalyahu A, et al. 2011. Absence of CNTNAP2 leads to epilepsy, neuronal migration abnormalities, and core autism-related deficits. *Cell* 147:235–46
95. Pieretti M, Zhang FP, Fu YH, Warren ST, Oostra BA, et al. 1991. Absence of expression of the *FMR-1* gene in fragile X syndrome. *Cell* 66:817–22
96. Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, et al. 2010. Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* 466:368–72
97. Ramocki MB, Zoghbi HY. 2008. Failure of neuronal homeostasis results in common neuropsychiatric phenotypes. *Nature* 455:912–18

98. Ronald A, Larsson H, Anckarsater H, Lichtenstein P. 2010. A twin study of autism symptoms in Sweden. *Mol. Psychiatry* 16:1039–47
99. Rujescu D, Ingason A, Cichon S, Pietiläinen OP, Barnes MR, et al. 2009. Disruption of the neurexin 1 gene is associated with schizophrenia. *Hum. Mol. Genet.* 18:988–96
100. Sanders SJ, Ercan-Sencicek AG, Hus V, Luo R, Murtha MT, et al. 2011. Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron* 70:863–85
101. Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, et al. 2012. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* 485:237–41
102. Sato D, Lionel AC, Leblond CS, Prasad A, Pinto D, et al. 2012. *SHANK1* deletions in males with autism spectrum disorder. *Am. J. Hum. Genet.* 90:879–87
103. Schenck A, Bardoni B, Langmann C, Harden N, Mandel JL, Giangrande A. 2003. CYFIP/Sra-1 controls neuronal connectivity in *Drosophila* and links the Rac1 GTPase pathway to the fragile X protein. *Neuron* 38:887–98
104. Schmeisser MJ, Ey E, Wegener S, Bockmann J, Stempel AV, et al. 2012. Autistic-like behaviours and hyperactivity in mice lacking ProSAP1/Shank2. *Nature* 486:256–60
105. Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, et al. 2007. Strong association of de novo copy number mutations with autism. *Science* 316:445–49
106. Sharp AJ, Mefford HC, Li K, Baker C, Skinner C, et al. 2008. A recurrent 15q13.3 microdeletion syndrome associated with mental retardation and seizures. *Nat. Genet.* 40:322–28
107. Shinawi M, Schaaf CP, Bhatt SS, Xia Z, Patel A, et al. 2009. A small recurrent deletion within 15q13.3 is associated with a range of neurodevelopmental phenotypes. *Nat. Genet.* 41:1269–71
108. Silverman JL, Yang M, Lord C, Crawley JN. 2010. Behavioural phenotyping assays for mouse models of autism. *Nat. Rev. Neurosci.* 11:490–502
109. Skafidas E, Testa R, Zantomio D, Chana G, Everall IP, Pantelis C. 2013. Predicting the diagnosis of autism spectrum disorder using gene pathway analysis. *Mol. Psychiatry.* In press. doi: 10.1038/mp.2012.126
110. Somerville MJ, Mervis CB, Young EJ, Seo EJ, del Campo M, et al. 2005. Severe expressive-language delay related to duplication of the Williams-Beuren locus. *N. Engl. J. Med.* 353:1694–701
111. Stefansson H, Rujescu D, Cichon S, Pietiläinen OP, Ingason A, et al. 2008. Large recurrent microdeletions associated with schizophrenia. *Nature* 455:232–36
112. Sutcliffe JS, Jiang YH, Galijaard RJ, Matsuura T, Fang P, et al. 1997. The E6-Ap ubiquitin-protein ligase (*UBE3A*) gene is localized within a narrowed Angelman syndrome critical region. *Genome Res.* 7:368–77
113. Szafranski P, Schaaf CP, Person RE, Gibson IB, Xia Z, et al. 2010. Structures and molecular mechanisms for common 15q13.3 microduplications involving *CHRNA7*: benign or pathological? *Hum. Mutat.* 31:840–50
114. Szatmari P, Paterson AD, Zwaigenbaum L, Roberts W, Brian J, et al. 2007. Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nat. Genet.* 39:319–28
115. Tabuchi K, Blundell J, Etherton MR, Hammer RE, Liu X, et al. 2007. A neuroligin-3 mutation implicated in autism increases inhibitory synaptic transmission in mice. *Science* 318:71–76
116. Toro R, Konyukh M, Delorme R, Leblond C, Chaste P, et al. 2010. Key role for gene dosage and synaptic homeostasis in autism spectrum disorders. *Trends Genet.* 26:363–72
117. van Bon BW, Mefford HC, Menten B, Koolen DA, Sharp AJ, et al. 2009. Further delineation of the 15q13 microdeletion and duplication syndromes: a clinical spectrum varying from non-pathogenic to a severe outcome. *J. Med. Genet.* 46:511–23
118. Van der Aa N, Rooms L, Vandeweyer G, van den Ende J, Reyniers E, et al. 2009. Fourteen new cases contribute to the characterization of the 7q11.23 microduplication syndrome. *Eur. J. Med. Genet.* 52:94–100
119. van der Zwaag B, Staal WG, Hochstenbach R, Poot M, Spierenburg HA, et al. 2010. A co-segregating microduplication of chromosome 15q11.2 pinpoints two risk genes for autism spectrum disorder. *Am. J. Med. Genet. B* 153B:960–66
120. Veltman MW, Craig EE, Bolton PF. 2005. Autism spectrum disorders in Prader-Willi and Angelman syndromes: a systematic review. *Psychiatr. Genet.* 15:243–54

121. Verhoeven WM, Egger JI, Willemsen MH, de Leijer GJ, Kleefstra T. 2012. Phelan-McDermid syndrome in two adult brothers: atypical bipolar disorder as its psychopathological phenotype? *Neuropsychiatr. Dis. Treat.* 8:175–79
122. Voineagu I, Wang X, Johnston P, Lowe JK, Tian Y, et al. 2011. Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* 474:380–84
123. Vorstman JA, Staal WG, van Daalen E, van Engeland H, Hochstenbach PF, Franke L. 2006. Identification of novel autism candidate regions through analysis of reported cytogenetic abnormalities associated with autism. *Mol. Psychiatry* 11:18–28
124. Vorstman JA, van Daalen E, Jalali GR, Schmidt ER, Pasterkamp RJ, et al. 2010. A double hit implicates DIAPH3 as an autism risk gene. *Mol. Psychiatry* 16:442–51
125. Vucurovic K, Landais E, Delahaigue C, Eutrope J, Schneider A, et al. 2012. Bipolar affective disorder and early dementia onset in a male patient with *SHANK3* deletion. *Eur. J. Med. Genet.* 55:625–29
126. Waddington CH. 1942. Canalization of development and the inheritance of acquired characters. *Nature* 150:563–65
127. Wang K, Zhang H, Ma D, Bucan M, Glessner JT, et al. 2009. Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature* 459:528–33
128. Weiss LA, Arking DE, Daly MJ, Chakravarti A. 2009. A genome-wide linkage and association scan reveals novel loci for autism. *Nature* 461:802–8
129. Weiss LA, Shen Y, Korn JM, Arking DE, Miller DT, et al. 2008. Association between microdeletion and microduplication at 16p11.2 and autism. *N. Engl. J. Med.* 358:667–75
130. Woltereck R. 1909. Weitere experimentelle Untersuchungen über Artveränderung, speziel über das Wesen quantitativer Artunterschiede bei Daphniden. *Verh. Dtsch. Zool. Ges.* 19:110–73
131. Won H, Lee HR, Gee HY, Mah W, Kim JI, et al. 2012. Autistic-like social behaviour in *Shank2*-mutant mice improved by restoring NMDA receptor function. *Nature* 486:261–65
132. Yamasaki K, Joh K, Ohta T, Masuzaki H, Ishimaru T, et al. 2003. Neurons but not glial cells show reciprocal imprinting of sense and antisense transcripts of *Ube3a*. *Hum. Mol. Genet.* 12:837–47
133. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, et al. 2010. Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* 42:565–69
134. Yaspan BL, Bush WS, Torstenson ES, Ma D, Pericak-Vance MA, et al. 2011. Genetic analysis of biological pathway data through genomic randomization. *Hum. Genet.* 129:563–71
135. Yu TW, Chahrour MH, Coulter ME, Jiralerspong S, Okamura-Ikeda K, et al. 2013. Using whole-exome sequencing to identify inherited causes of autism. *Neuron* 77:259–73
136. Zweier C, de Jong EK, Zweier M, Orrico A, Ousager LB, et al. 2009. *CNTNAP2* and *NRXN1* are mutated in autosomal-recessive Pitt-Hopkins-like mental retardation and determine the level of a common synaptic protein in *Drosophila*. *Am. J. Hum. Genet.* 85:655–66

Annual Review of Genomics and Human Genetics

Volume 14, 2013

Contents

Errata

An online log of corrections to *Annual Review of Genomics and Human Genetics* articles may be found at http://genom.annualreviews.org

Citation: Transl Psychiatry (2012) 2, e179; doi:10.1038/tp.2012.102

www.nature.com/tp

Analysis of the chromosome X exome in patients with autism spectrum disorders identified novel candidate genes, including *TMLHE*

C Nava[1,2,3,4], F Lamari[5,28], D Héron[4,6,7,8,28], C Mignot[4,6,7,8], A Rastetter[1,2,3], B Keren[9], D Cohen[10,11], A Faudet[4], D Bouteiller[1,2,3,12], M Gilleron[5], A Jacquette[4,7,8], S Whalen[4], A Afenjar[4,6,7,8,13], D Périsse[10,14], C Laurent[1,2,10], C Dupuits[1,15], C Gautier[1,2], M Gérard[16], G Huguet[17,18,19], S Caillet[4,15], B Leheup[20], M Leboyer[21,22,23,24], C Gillberg[25], R Delorme[26], T Bourgeron[17,18,19], A Brice[1,2,3,4] and C Depienne[1,2,3,27]

The striking excess of affected males in autism spectrum disorders (ASD) suggests that genes located on chromosome X contribute to the etiology of these disorders. To identify new X-linked genes associated with ASD, we analyzed the entire chromosome X exome by next-generation sequencing in 12 unrelated families with two affected males. Thirty-six possibly deleterious variants in 33 candidate genes were found, including *PHF8* and *HUWE1*, previously implicated in intellectual disability (ID). A nonsense mutation in *TMLHE*, which encodes the ε-*N*-trimethyllysine hydroxylase catalyzing the first step of carnitine biosynthesis, was identified in two brothers with autism and ID. By screening the *TMLHE* coding sequence in 501 male patients with ASD, we identified two additional missense substitutions not found in controls and not reported in databases. Functional analyses confirmed that the mutations were associated with a loss-of-function and led to an increase in trimethyllysine, the precursor of carnitine biosynthesis, in the plasma of patients. This study supports the hypothesis that rare variants on the X chromosome are involved in the etiology of ASD and contribute to the sex-ratio disequilibrium.
Translational Psychiatry (2012) **2,** e179; doi:10.1038/tp.2012.102; published online 23 October 2012

Introduction

Autism spectrum disorders (ASD) constitute a common but heterogeneous group of neurodevelopmental disorders characterized by impairment of social interactions and communication, stereotyped behaviors and restricted interests. Although they are probably the most heritable of psychiatric conditions, with a concordance rate of ~80–90% in monozygotic twins versus ~10–20% in dizygotic twins, few genes have been reliably associated with ASD.[1–3] Recent studies have highlighted the vast heterogeneity and complexity of the genetics of these disorders. All mutations or copy number variants (CNVs) associated so far with ASD have been rare, with minor allele frequencies <1%. A few *de novo* or inherited CNVs, some of which are recurrent, such as duplications of the 15q11-q13 or 7q11.23 and deletions of 16p11.2 regions, were shown to confer a highly penetrant risk of autism.[4–7] More recently, *de novo* mutations in various highly interconnected genes were shown to contribute to ASD, suggesting that abnormalities in different genes could converge to alter common pathways.[8–10] Abnormalities in at least two pathways were repeatedly related to ASD: the first includes

[1]INSERM, U975—CRICM, Institut du cerveau et de la moelle épinière (ICM), Hôpital Pitié-Salpêtrière, Paris, France; [2]CNRS 7225—CRICM, Hôpital Pitié-Salpêtrière, Paris, France; [3]Université Pierre et Marie Curie-Paris-6 (UPMC), UMR_S 975, Paris, France; [4]Département de Génétique et de Cytogénétique, Unité fonctionnelle de génétique clinique, AP-HP, Hôpital Pitié-Salpêtrière, Paris, France; [5]Département de Biochimie, AP-HP, Hôpital Pitié-Salpêtrière, Paris, France; [6]AP-HP, Hôpital Trousseau, service de neuropédiatrie, Paris, France; [7]Centre de Référence 'déficiences intellectuelles de causes rares', Paris, France; [8]Groupe de Recherche Clinique (GRC) 'déficience intellectuelle et autisme' UPMC, Paris, France; [9]Département de Génétique et de Cytogénétique, Unité fonctionnelle de cytogénétique, AP-HP, Hôpital Pitié-Salpêtrière, Paris, France; [10]Service de psychiatrie de l'enfant et de l'adolescent, AP-HP, Hôpital Pitié-Salpêtrière, Paris, France; [11]Institut des Systèmes Intelligents et Robotiques, CNRS UMR 7222, UPMC-Paris-6, Paris, France; [12]ICM, PFGS Platform, Hôpital Pitié-Salpêtrière, Paris, France; [13]Centre de référence des anomalies du développement et syndromes malformatifs, Hôpital Trousseau, Paris, France; [14]Centre référent autisme, Paris, France; [15]Service de diététique et unité fonctionnelle de neurométabolisme, AP-HP, Hôpital Pitié-Salpêtrière, Paris, France; [16]CHU Côte de Nacre, Paris, France; [17]Institut Pasteur, Human Genetics and Cognitive Functions Unit, Paris, France; [18]CNRS URA 2182 'Genes, synapses and cognition', Institut Pasteur, Paris, France; [19]University Paris Diderot, Sorbonne Paris Cité, Human Genetics and Cognitive Functions, Paris, France; [20]CHU de Nancy Pôle Enfants, Service de Médecine Infantile III et Génétique Clinique, Centre de référence Anomalies du développement et Syndromes malformatifs et Université de Lorraine EA 4368, Vandoeuvre les Nancy, France; [21]Inserm, U955, Créteil, France; [22]Université Paris Est, Faculté de médecine, Créteil, France; [23]AP-HP, Hôpital H. Mondor—A. Chenevier, Pole de Psychiatrie, Créteil, France; [24]Fondation FondaMental, Créteil, France; [25]Department of Child and Adolescent Psychiatry, Goteborg University, Goteborg, Sweden; [26]AP-HP, Hôpital Robert Debré, Service de pédopsychiatrie, Paris, France and [27]Département de Génétique et de Cytogénétique, Unité fonctionnelle de neurogénétique moléculaire et cellulaire, AP-HP, Hôpital Pitié-Salpêtrière, Paris, France

Correspondence: Dr C Depienne, or Professor A Brice, INSERM U975 (Cricm), Institut du cerveau et de la moelle épinière, Hôpital Pitié-Salpêtrière, Paris 75 013, France.

E-mail: christel.depienne@upmc.fr or alexis.brice@upmc.fr

[28]These authors contributed equally to the study.

Keywords: Autism spectrum disorders; carnitine; chromosome X; male excess; next-generation sequencing; *TMLHE*

Received 20 August 2012; accepted 20 August 2012

mutations in *TSC1/TSC2*, *NF1* or *PTEN* in the mTOR (mammalian target of rapamycin) pathway; the second is illustrated by mutations in *NLGN3–4*, *SHANK1–3* and *NRXN1*, all of which encode synaptic proteins.[11–13]

A striking feature of ASD is the excess of affected males, with a sex-ratio disequilibrium of 4:1 that reaches 10:1 in patients with normal cognitive abilities.[14] This suggests that genes located on sex chromosomes contribute to the etiology of ASD or that the penetrance of autistic traits depends on sex determinants such as hormones. In favor of the first hypothesis, mutations in *NLGN4X* and *NLGN3* on chromosome X have been identified in a few families with ASD.[15] Additionally, the analysis of all or selected genes located on the X chromosome successfully identified new candidate genes for intellectual disability (ID),[16] ASD and schizophrenia.[17] Interestingly, the risk of recurrence of ASD is significantly increased in families with two affected sibs, reaching 32% and more when both affected subjects are males.[18] This suggests that highly penetrant forms of ASD with autosomal recessive or X-linked inheritance have been overlooked.

To test the hypothesis that yet undiscovered X-linked genes are associated with highly penetrant forms of ASD, we selected 12 unrelated families with at least two affected males compatible with X-linked inheritance and analyzed all of the coding regions on the X chromosome.

Material and methods

Patients. The entire exome of chromosome X was sequenced in 12 families with two affected males with ASD or ID compatible with X-linked inheritance, recruited from the 'Centre de Référence Déficiences Intellectuelles de causes rares' (Pitié-Salpêtrière Hospital; Supplementary Figure S1). Index cases were evaluated by specialized geneticists and pediatric neurologists and/or child psychiatrists. Patients were assessed with the Autism Diagnostic Interview-Revised. Nine index cases had autism with ID and three had Asperger syndrome or high-functioning autism based on DSM IV-TR (Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision) criteria. Clinical features of the index cases and the affected relatives are detailed in Supplementary Clinical Table. Normal results were previously obtained by karyotyping, searches for fragile-X syndrome, microarray analysis (CytoSNP-12, Illumina, San Diego, CA, USA) and sequencing of *NLGN3–4X* and *SHANK3* when appropriate, as well as metabolic screening (including at least creatine and guanidinoacetate analysis).

For *TMLHE* screening, a cohort of 161 patients (134 patients with autism and ID and 27 patients with Asperger syndrome) recruited at the Pitié-Salpêtrière Hospital (Centre de référence Déficiences Intellectuelles de causes rares or Centre référent diagnostic autisme, Paris, France) and 340 patients from the PARIS (Paris Autism Research International sib pair) study (including 194 patients with autism and ID and 59 patients with Asperger syndrome) were included, for a total of 501 unrelated male patients with ASD. In addition, 765 healthy male controls from North Africa ($n=320$), Europe ($n=350$) and Lebanon ($n=95$) were included to test the new variants.

The analysis of microrearrangements in *TMLHE* included 178 additional patients with ASD from the PARIS studies previously included in Autism Genome Project (AGP, http://www.autismgenome.org/)[7] and 896 healthy male individuals. The control groups included 371 European male subjects from La Pitié-Salpêtrière hospital, 142 from other European laboratories and 383 control individuals from the Study on Addiction Genetics and Environment ($n=371$) and HapMap CEPH Utah ($n=12$) series.[7] Raw intensities and genotypes were obtained from NHGRI-dbGaP (http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000092.v1.p1). The sub-set of control data set used in the specific CNV analyses in this paper is composed of control samples that passed all quality control filters (Log R ratios s.d. $=0.27$; B allele frequency s.d. $=0.13$; Call Rate >0.99). Informed written consent was obtained from each individual or his/her parents before blood sampling. All experiments were performed in accordance with French guidelines and rules.

Next-generation sequencing. Next-generation sequencing was performed at Integragen SA (Evry, France). Regions of the X chromosome corresponding to coding and 3′–5′ untranslated region (UTR) sequences were captured from genomic DNA using a custom Agilent SureSelect Target Enrichment methodology (Agilent, Santa Clara, CA, USA) with the biotinylated oligonucleotide probe library, followed by paired-end 75 b massively parallel sequencing on Illumina GAIIx (Illumina). For detailed explanations of the process, see Gnirke *et al.*[19] Sequence capture, enrichment and elution were performed precisely according to the manufacturer's instructions and protocols. Briefly, 3 μg of each genomic DNA was fragmented by sonication to yield fragments of 150–200 bp and then purified. Paired-end adapter oligonucleotides from Illumina were ligated on repaired A-tailed fragments, then purified and enriched by six PCR cycles. Then 500 ng of these purified libraries were hybridized to the SureSelect oligo probe capture library for 24 h. After hybridization, washing and elution, the eluted fraction was PCR-amplified for 10–12 cycles, purified and quantified by quantitative PCR to obtain sufficient DNA template for downstream applications. Each eluted-enriched DNA sample was sequenced on an Illumina GAIIx as paired-end 75 b reads. Image analysis and base calling was performed using Real Time Analysis Pipeline version 1.9 (Illumina) with default parameters.

Bioinformatics analysis. Sequencing data was analyzed according to the Illumina pipeline (CASAVA1.7) and aligned with the Human reference genome (Hg19) using the ELANDv2 algorithm. Variant annotation (RefSeq gene annotation) identification of known polymorphisms (referenced in dbSNP or 1000 Genome) and analysis of the position and consequences of the variants (for example, exonic, intronic, silent and nonsense), was determined with an in-house pipeline from the positions included in the bait coordinates. The frequencies (in the homozygous or heterozygous state) were determined from all exomes sequenced at Integragen and from exome results provided by HapMap. Results per sample were obtained in tabulated text files, and coverage/depth statistical analyses were performed for each

bait. The 13,464 single nucleotide polymorphisms (SNPs) and 1532 indels in the 12 male index cases included 1467 SNPs (10.9%) and 590 indels (38.5%) predicted to be at the heterozygous state. Eleven of these variants were tested by Sanger sequencing and proved to be false-positives. Further analysis focused, therefore, on variants predicted to be hemizygous. A total of 171 hemizygous variants were tested and confirmed by Sanger sequencing. The strategy used for selecting potentially pathogenic variants is detailed in Figure 1. They: (i) were located in chromosome X regions common to the two affected sibs (4331 SNPs and 299 indels in 11 families); (ii) had with a minor allele frequency $<1\%$ in dbSNP135 (http://www.ncbi.nlm.nih.gov/projects/SNP/), Exome variant server (http://evs.gs.washington.edu/EVS/) and 29 other exomes; (iii) were found in genes expressed in brain according to the Unigene (http://www.ncbi.nlm.nih.gov/unigene) or Uniprot (http://www.uniprot.org/) databases; and (iv) were predicted to have an impact on the gene or the protein (nonsense variants, missense variants, predicted at least once *in silico* to be deleterious, and synonymous, intronic or 5–3′UTR variants with possible effects on splice sites or promoters using Alamutv2.1/AlamutHT). For variants present in at least two index cases, only those segregating in all the affected members of all families were retained. Mutation interpretation and amino-acid conservation in orthologs and paralogs were assessed using the Alamutv2.1/AlamutHT softwares (Interactive Biosoftware, Rouen, France). Prediction of pathogenicity was assessed using PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/), SIFT (scale-invariant feature transform) (http://sift.bii.a-star.edu.sg/), Mutpred (http://mutpred.mutdb.org/) and SNPs&GO (http://snps-and-go.biocomp.unibo.it/snps-and-go/). Frequencies were compared with the Fisher's exact test.

High-density SNP arrays. Index cases and affected relatives were screened using Illumina cytoSNP-12 arrays to search for CNVs and identify regions on the X chromosome shared by the affected sibs. Illumina microarray experiments were automated and performed at the P3S platform (Pitié-Salpêtrière Hospital), according to the manufacturer's specifications (Illumina, San Diego, CA, USA). Image acquisition was performed using a BeadArray Reader (Illumina). Image data analysis and automated genotype calling was performed using GenomeStudiov2011.1 (Illumina). Genomic positions were based on the UCSC and Ensembl Genome Browsers. Genotypes on chromosome X were analyzed in affected relatives of each family to identify shared X regions (with the exception of family 12 for which SNP microarray data were unavailable for the affected uncle). Shared regions were defined as identical genotypes spanning at least 2 Mb.

For the analysis of *TMLHE* microrearrangements, control individuals ($n=525$) and ASD cases ($n=356$) were genotyped using Illumina Human 1M-single BeadChip arrays. Samples were processed using the manufacturer's recommended protocol, and BeadChips were scanned on the Illumina BeadArray Reader using default settings. Analysis and intra-chip normalization were performed using Illumina's BeadStudio software v.3.3.7, with a GenCall cutoff of 0.1. The quality-control criteria were selected: the Array with call rate $>95\%$; standard deviation for log R ratio values in the autosomes <0.35; and standard deviation of the B Allele frequency values (that is, allelic ratios within the 0.25–0.75 ranges) >0.13. For the samples that passed the above SNP and intensity quality-control filters, we used the QuantiSNP[20] and visualizator SnipPeep CNV calling algorithms. The required data for CNV analysis, that is, within-sample normalized fluorescence (that is, X and Y normalized values), between-sample normalized fluorescence (that is, Log R ratios and B allele frequency values) and genotypes for each sample, were exported directly from Illumina's Beadstudio software. We excluded CNVs when they failed stringent quality control criteria: <5 consecutive probes covering 1 kb of sequence were merged using outside probe boundaries (that is, union of the CNVs) and low confidence score log Bayes factor <15.

Sanger sequencing. Specific primer pairs were designed to amplify 182 variants detected by next-generation sequencing. In addition, eight primer pairs were designed to amplify the coding exons and adjacent intron–exon boundaries of the *TMLHE* gene. Primer sequences are provided in supplemental data. Forward and reverse sequence reactions were performed with the Big Dye Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems, Life Technologies Corporation, Carlsbad, CA, USA) using the same primers. G50-purified sequence products were run on an ABI 3730 automated sequencer (Applied Biosystems) and data were analyzed with Seqscape v2.6 software (Applied Biosystems). The mutation nomenclature is based on the *TMLHE* cDNA reference sequence (NM_018196.3).

Quantitative multiplex PCR and long-range PCR. The presence of a deletion of exon 2 in *TMLHE* was tested in the 501 patients with ASD and 371 male control individuals by quantitative multiplex PCR. One hundred and seventy-eight patients with ASD were also screened using Illumina Human 1 M-single BeadChip arrays with the same results. Two primer pairs were used in the quantitative multiplex assay: one specific of exon 2 of *TMLHE* (final concentration: 0.5 μM) and one amplifying exon 2 of GPR128 (final concentration: 0.08 μM). PCR conditions were as follows: 96 °C 5 min, 20 cycles: 94 °C 30 s, 60 °C − 0.5 °C/cycle 30 s, 72 °C 40 s, and 15 cycles: 94 °C 30 s, 50 °C 30 s, 72 °C 40 s, followed by 7 min at 72 °C. PCR products were quantified on a Caliper LabChip system (Caliper Life Sciences, Hopkinton, MA, USA). In addition, the presence of the deletion of exon 2 in *TMLHE* was confirmed by long-range PCR using the SequalPrep Long PCR kit (Invitrogen, Life Technologies Corporation, Carlsbad, CA, USA) according to the manufacturer's recommendations. The PCR conditions were as follows: 2 min at 94 °C, 10 cycles: 94 °C 10 s, 58 °C 30 s, 68 °C 18 min (1 min kb^{-1}), 25 cycles: of 94 °C 10 s, 58 °C 30 s, 68 °C 18 min (1 min kb^{-1}) + 20 s/cycle, followed by 5 min at 72 °C. Frequencies were compared with the Fisher's exact test.

Cell culture and mRNA experiments. Lymphoblasts from the two affected brothers and the mother of family 9 were isolated from peripheral blood cells using standard procedures. Fibroblasts were taken from skin cells of the affected

brothers. Lymphoblastic cells and fibroblasts were pre-treated, or not, overnight with 10 μg ml^{-1} emetin, an inhibitor of nonsense-mediated decay.

Total RNA from lymphoblasts and fibroblasts was isolated using the Qiagen RNeasy Mini kit (Invitrogen). cDNAs were synthesized from 1 μg of total RNA using the SuperScript III First-Strand Kit (Invitrogen). The reverse-transcribed *TMLHE* cDNA was amplified and sequenced using specific primers located in exons 2 (forward) and 4 (reverse). The PCR products were run on 2% agarose gels. *TMLHE* mRNA was quantified using the Qiagen QuantiTect primer assays for *TMLHE* (forward and reverse primers located in exons 7 and 8 of *TMLHE*). *PPIA* was used as the control gene. Each sample was run in triplicate on a Lightcycler 480 (Roche, Applied Sciences, Penzberg, Germany). Forty-five two-step cycles (15 s at 95 °C and 30 s at 60 °C) were performed and analyzed using Lightcycler 480 software release 1.5.0. Relative abundance was calculated using the formula $r = 2 - \Delta\Delta Ct$, where $\Delta\Delta Ct = (Ct_{\text{Gene tested}} - Ct_{\text{control genes}})_{TMLHE} - (Ct_{\text{Gene tested}} - Ct_{\text{control genes}})_{PPIA}$.

Chromatography and mass spectrometry. A mixture of internal standard solution was prepared by dissolving 30 μM of carnitine-(*N*-methyl-d_3) (Cambridge Isotopes Lab, Andover, MA, USA) and 15 μM of ε-N-trimethyl-($^{13}C_3$)-L-Lysine (Sigma-Aldrich, St Louis, MO, USA) in methanol. In all, 20 μl of internal standard mixture was added to 30 μl of plasma or urine after mixing, 100 μl of methanol was added and mixed with vortex for protein precipitation. The mixture was incubated in ice for 15 min and centrifuged at 15000 *g* for 10 minutes at +4 °C. The supernatant was transferred into the vials and 5 μl were injected into tandem LCMS/MS system. Calibration curves were performed by serial dilution of stock solution containing 46.8 μM of l-carnitine and 20 μM of ε-*N*-trimethyllysine (Sigma-Aldrich) in methanol. An Acquity UPLC (ultra performance liquid chromatography) chromatographic system equipped with a BEH C18 RP column (1.7 μ, 50 mm × 2) maintained at 45 °C was coupled to a TQD (tandem quadrupole detector) MS/MS system (Waters, Guyancourt, France) and used as an LCMS/MS (liquid chromatography–mass spectrometry/mass spectrometry) system for trimethyllysine (TML) and carnitine measurement. The mobile phases were: eluent A, ultrapure water; eluent B, acetonitrile. The elution gradient was as follows: flow rate 0.8 ml min^{-1}, 0–1 min, 0% A; 1–1.2, 0–100% A; 1.2–2 min, 100%; 2–2.2, 100–0% A; 2.2–5, column equilibration with 100% B. The detector was used in multiple-reaction monitoring to detect the transition of a specific precursor to daughter ions 189.1/84.1 and 162.2/103.1 for N^6-trimethyllysine and carnitine, respectively.

Dietary assessment. In order to determine prospective and retrospective dietary intakes, the patient and their parents had an interview with a dietician. A dietary questionnaire listing all the food and beverages consumed by the patient for a period of 3 days was completed by the parents. Nutritional intakes were estimated with the DSMS software. Carnitine intakes were calculated based on the reference table of the Linus Pauling Institute Oregon State University.

Results

Chromosome X exome sequencing. The coding regions of all genes on chromosome X, including 5′–3′ UTRs, were sequenced in 12 index cases using next-generation sequencing (pedigrees are shown in Supplementary Figure S1). A mean number of 1000 SNPs (829–1459) and 78 indels (56–113; Table 1 and Supplementary Table SA and SB) were identified per patient. In parallel, chromosome X regions common to the affected sibs, ranging from 15 to 109 Mb, were identified in 11 families using Illumina cytoSNP-12 arrays (Supplementary Figure S2). The mean numbers of SNPs and indels located in shared X regions were 394 (77–828) and 27 (7–77) per patient, respectively. Further analysis focused on variants that were (i) absent or rare in databases (minor allele frequency <1%), (ii) located in genes expressed in brain and (iii) predicted to be deleterious (Figure 1). Thirty-eight possibly deleterious variants

Table 1 Summary of all SNPs and indels detected by exome analysis of chromosome X in the index cases of families 1–12 and those located in chromosome X regions shared by families 1–11

Family	*Nb SNPs (novel)*	*Nb indels (novel)*	*Linked regions (Mb)*	*SNPs in linked regions*					*Indels in linked regions*		
				Total	*Nonsense (novel)*	*Missense (novel)*	*Synonymous (novel)*	*Non-coding (novel)*	*Total*	*Exons (novel)*	*Non-coding (novel)*
Family 1	1233 (54)	108 (45)	**109**	**828**	1 (0)	122 (3)	137 (3)	568 (23)	**77**	5 (1)	72 (30)
Family 2	1039 (40)	82 (37)	**37**	**369**	1 (0)	59 (3)	48 (1)	261 (8)	**19**	0	19 (6)
Family 3	871 (16)	56 (22)	**33**	**251**	1 (0)	36 (0)	40 (1)	174 (2)	**15**	0	15 (9)
Family 4	861 (27)	67 (23)	**57**	**340**	1 (0)	72 (4)	58 (1)	209 (2)	**21**	2 (1)	19 (8)
Family 5	892 (30)	81 (34)	**76**	**412**	1 (0)	57 (1)	66 (0)	288 (15)	**36**	2 (1)	34 (14)
Family 6	989 (33)	83 (37)	**27**	**319**	0	54 (2)	46 (2)	219 (7)	**28**	3 (2)	25 (14)
Family 7	829 (18)	66 (27)	**36**	**200**	1 (0)	37 (1)	41 (0)	121 (2)	**14**	2 (1)	12 (6)
Family 8	1459 (63)	113 (60)	**42**	**614**	1 (0)	71 (3)	101 (1)	441 (22)	**42**	4 (3)	38 (24)
Family 9	1098 (35)	82 (33)	**81**	**656**	1 (1)	93 (2)	90 (2)	472 (21)	**23**	1 (0)	22 (7)
Family 10	835 (31)	63 (24)	**15**	**77**	0	6 (0)	10 (1)	61 (2)	**7**	0	7 (1)
Family 11	905 (24)	73 (32)	**53**	**265**	3 (0)	38 (1)	38 (1)	186 (6)	**17**	1 (1)	16 (5)
Family 12	986 (24)	68 (19)	**Unknown**								

Abbreviation: SNP, single-nucleotide polymorphism.
Linked regions and the total number of SNPs and indels in linked regions are indicated in bold.

(mean number per family: 3.2 range: 0–9), all confirmed by Sanger sequencing, were detected (Table 2). Analysis of matched control populations excluded two variants that had a frequency $\geqslant$1%. The variants were present in 15/22 asymptomatic male relatives. Altogether, these results identified 36 rare, possibly deleterious variants in 33 different genes in 9 families.

In two families, the variants were in genes previously implicated in ID.[21–24] Both variants, c.2904_2906del/p.Ser969del in *PHF8* and c.2849T>A/p.Val950Asp in *HUWE1*, affect amino-acids that were highly conserved during evolution and were not found in a large control population or reported in Hapmap, 1000 Genomes and the Exome Variant Server (Figure 2). Although, mutations in *PHF8* causing a loss-of-function were previously identified in patients with ID and cleft lip/palate,[21–23] the p.Ser969del variant segregated with high-functioning autism without other clinical features in family 8. In the index case of family 4, p.Val950Asp in *HUWE1* was predicted to be deleterious by SIFT, Polyphen-2 and Mutpred algorithms. Surprisingly, this variant was not found in the proband's brother, who was less severely affected, and turned out to have occurred *de novo* in the proband. We hypothesized that Val950Asp in *HUWE1* contributed to a genetically complex disorder by increasing the severity of the phenotype. Alternatively, the phenotype of the brother could have had a different etiology. These results illustrate the complexity of inheritance in ASD, in which a combination of rare inherited and *de novo* events can contribute to the disorder.[6–10]

Among the remaining variants, a nonsense mutation (c.229C>T/p.Arg77X) in *TMLHE*, encoding ε-*N*-trimethyllysine hydroxylase, the enzyme catalyzing the first step of carnitine biosynthesis from TML, segregated with autism and moderate ID in family 9 (Figures 3a and b). The p.Arg77X mutation was absent from 508 healthy male controls and databases.

This study identified several additional rare variants that might contribute to ASD such as c.521C>A/p.Ala174Asp (family 1) in *ODZ1*, which encodes teneurin-1, a transmembrane protein expressed in the developing central nervous system that might have a role in neuronal connectivity.[25] In addition, 11 variants located in introns or UTRs of genes expressed in the brain and predicted to have a possible effect on gene expression were found in five families. Five of these genes are involved in axon guidance (*PLXNA3*, *PLXNB3*, *KAL1*) or neurotransmission (*SYN1*, *GABRE*), four regulate transcription, splicing or translation (*TXLNG*, *TSPYL2*, *AFF2*, involved in brain development, and *RBM3*, involved in RNA processing, regulation of translation and production of miRNA);[26] the remaining genes are involved in ubiquitination (*KLHL13*) or in protein transport (*BCAP31*).

Finally, no rare variants meeting the criteria defined in Material and methods were found in three families (families 3, 7 and 10), suggesting that genetic factors in these families are possibly located in unexplored regions of the X chromosome, on autosomes, or were present with a frequency $\geqslant$1%.

Screening of TMLHE and functional consequences of mutations. To investigate the effect of the p.Arg77X mutation at the mRNA level, we performed quantitative reverse transcriptase–PCR analysis in lymphoblasts and fibroblasts from the two affected brothers of family 9 and their mother. The mutated mRNA was significantly downregulated in the cells of the affected sibs. Pretreatment of the cells with emetin restored the mRNA levels, indicating that this downregulation corresponded to the degradation of the mutated mRNA by nonsense-mediated decay (Figure 3e). Interestingly, two mRNA isoforms were detectable in the

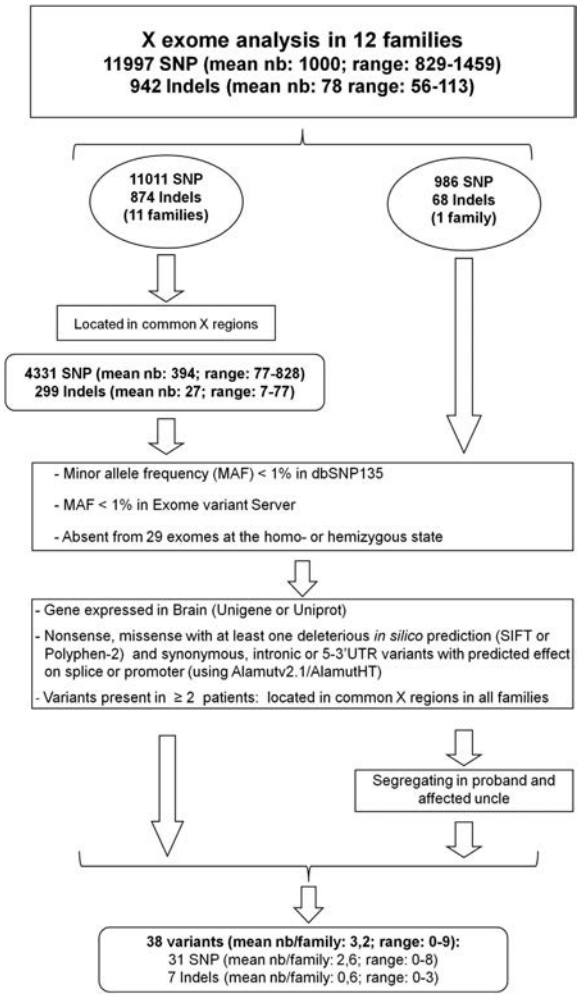


Figure 1 Strategy used for the selection of rare and possibly deleterious variants. Data from NGS and single nucleotide polymorphism (SNP) arrays were combined to conserve only variants located in X regions shared by the affected sibs (families 1–11). Further filters included a minor allele frequency (MAF) <1%, expression of the corresponding genes in brain and *in silico* predictions compatible with an effect of the variant on the gene or the protein (nonsense variants, missense variants with at least one prediction *in silico* by SIFT (scale-invariant feature transform) or Polyphen-2 that it is deleterious and synonymous, intronic or 5–3′UTR variants with possible effects on splice sites or promoters using Alamutv2.1/AlamutHT). For variants present in at least two index cases, only those that segregated in all affected members of all families were conserved. For one family (family 12), microarray data were unavailable for the affected uncle; segregation of variants found in the index case was performed at a later time.

Table 2 Summary of rare, possibly deleterious variants located in shared regions on chromosome X

Index case	*Position (Hg19)*	*Gene*	*NM_number*	*Nucleotide change*	*Amino-acid change*	*Mutation type*	*rs number (dbSNP135)*	*rsMAF(dbSNP 135/ Exome variant server) (%)*	*Grantham distance*	*SIFT*	*PolyPhen2*	*Mutpred*	*SNPs&GO*	*Controls*	*Variant found in healthy man in the family*	*Causative status*
Family 1	3228257	MXRA5	NM_015419	c.7987G >T	p.Gly2663Trp	Missense	rs143264543	0/0.02	184	deleterious	damaging	0.575	neutral	0/93	NA	possible
	16860727	TXLNG	NM_018360	c.*838A >G		3'UTR	rs140169598	0/-							NA	unknown
	18910664	PHKA2	NM_000292	c.*939T >C		3'UTR	rs142176794	0/-							NA	unknown
	19037757	GPR64	NM_001079858	c.670C >T	p.Pro224Ser	Missense	rs140334931	0/0.34	74	tolerated	damaging	0.268	neutral		NA	possible
	23801307	SAT1	NM_002970	c.-162G >A		5'UTR	rs113458958	0.8/-							NA	unknown
	118250487	KIAA1210	NM_020721	c.622G >T	p.Val208Phe	Missense	—	-/0.18	50	deleterious	damaging	0.174	unknown		NA	possible
	124028159	ODZ1	NM_001163278	c.521C >A	p.Ala174Asp	Missense	rs139486546	0/0.1	126	deleterious	benign	0.396	neutral	0/159	NA	possible
	148081861	AFF2	NM_002025	c.*8999G >C		3'UTR	rs142697456	0/-							NA	unknown
Family 2	17750042	NHS	NM_198270	c.4351G >C	p.Ala1451Pro	Missense	—	-/-	27	tolerated	damaging	0.345	neutral	0/88	NA	possible
	130411844	IGSF1	NM_001170961	c.2321A >G	p.Glu774Gly	Missense	—	-/-	98	deleterious	damaging	0.484	unknown	0/96	NA	possible
	151092546	MAGEA4	NM_001011548	c.410T >A	p.Leu137Gln	Missense	—	-/-	113	deleterious	damaging	0.827	disease	0/96	NA	possible
	153734897	FAM3A	NM_021806	c.*245A >G		3'UTR	rs145895727	0/-							NA	unknown
Family 4	3240238	MXRA5	NM_015419	c.3488G >A	p.Arg1163His	Missense	rs139106444	0/0.36	29	deleterious	damaging	0.399	neutral	1/130 male	yes (grand-father)	unlikely
	49105970	CCDC22	NM_014008	c.1636G >A	p.Asp546Asn	Missense	rs147222955	0/0.048	23	tolerated	damaging	0.625	neutral	1/155 male	yes (grand-father)	unlikely
	49142473	PPP1R3F	NM_033215	c.1321G >A	p.Glu441Lys	Missense	—	-/-	56	deleterious	benign	0.153	neutral	0/384	yes (grand-father)	possible
	53630356	HUWE1	NM_031407	c.2849T >A	p.Val950Asp	Missense	—	-/-	152	deleterious	damaging	0.702	neutral	0/329 male	no (grand-father)	probable
	75649236	MAGEE1	NM_020932	c.913A >T	p.Ser305Cys	Missense	rs142080557	0/0.16	112	tolerated	damaging	0.309	neutral		yes (grand-father)	unlikely
Family 5	117251283	KLHL13	NM_033495	c.-132040G >A		5'UTR	—	-/-						0/71 male	NA	unknown
	138287146	FGF13	NM_004114	c.-493981G >A		5'UTR	rs185138825	0/-						3/73 male	NA	unlikely
Family 6	152960226	SLC6A8	NM_005629	c.1649C >G	p.Thr550Ser	Missense	—	-/0.01	58	tolerated	damaging	0.284	neutral	0/218 male	yes (grand-father)	possible
	153210180	RENBP	NM_002910	c.-134_-118delins17		5'UTR	—	-/-							yes (grand-father)	unknown
	153701511	PLXNA3	NM_017514	c.*483 G >A		3'UTR	rs41311390	0.3/-							yes (grand-father)	unknown
	48432832	RBM3	NM_006743	c.-143C >G		5'UTR	—	-/-							yes (grand-father)	unknown
	152989468	BCAP31	NM_001139457	c.-350_-349insCA		5'UTR ins	—	-/-							yes (grand-father)	unknown
Family 7	120009218	CT47B1	NM_001145718	c.291_293del	p.Glu99del	Inframe del	—	-/-						3/40 male	NA	unlikely
Family 8	3227525	MXRA5	NM_015419	c.*232 G >T		3'UTR	rs143753500	0/-							no (2 uncles)	unknown
	8499107	KAL1	NM_000216	c.*1929 A >G		3'UTR	rs6530183	0/-							yes (2 uncles)	unknown
	53117400	TSPYL2	NM_022117	c.*279 A >C		3'UTR	rs113822334	0.9/-							no (2 uncles)	unknown
	53966801	PHF8	NM_001184896	c.2904_2906del	p.Ser969del	Inframe del	—	—						0/189	no (2 uncles)	probable
	63412635	FAM123B	NM_152424	c.532C >T	p.Arg178Cys	Missense	—	–/0.01	180	deleterious	damaging	0.506	neutral		no (2 uncles)	unlikely
	153037579	PLXNB3	NM_005393	c.2739–58_2739–57ins28		Intronic ins	—	-/-							yes (2 uncles)	unlikely
	47466513	SYN1	NM_006950	c.435+27G >C		Intronic	—	–/0.26							no (2 uncles)	unlikely
	55027965	APEX2	NM_014481	c.158–5del		Intronic del	—	-/-							yes (2 uncles)	unlikely
	154284285	FUNDC2	NM_023934	c.*1338C >G		3'UTR	rs146616394	0/-							yes (2 uncles)	unknown
Family 9	153210180	RENBP	NM_002910	c.-134_-118delins17		5'UTR	—	-/-							yes (grand-father)	unknown
	154754246	TMLHE	NM_018196	c.229C >T	p.Arg77X	Nonsense	—	-/-						0/508 male, 1 htz/184 female	no (grand-father and uncle)	probable
Family 11	54956008	TRO	NM_001039705	c.2851 G >A	p.Gly951Ser	Missense	—	-/-	56	deleterious	unknown	0.174	neutral	0/138 male	yes (heathy brother)	possible
Family 12	151123143	GABRE	NM_004961	c.*30C >T		3'UTR	—	-/-							NA	unknown

Abbreviation: MAF, minor allele frequency; SIFT, scale-invariant feature transform; UTR, untranslated region.
This list includes variants that are rare or absent from databases (MAFs <1%), are found in genes expressed in brain and have a possible impact on gene or protein function (nonsense, missense, predicted to be deleterious by at least one bioinformatics algorithm or in an intron, 5'UTR or 3'UTR with a predicted effect on a splice site or promoter). Mutpred: score <0.5: benign; 0.5–0.75: possibly disease-associated; and >0.75: probably disease-associated.Grantham distance: conservative (0–50), moderately conservative (51–100), moderately radical (101–150) or radical (⩾151).

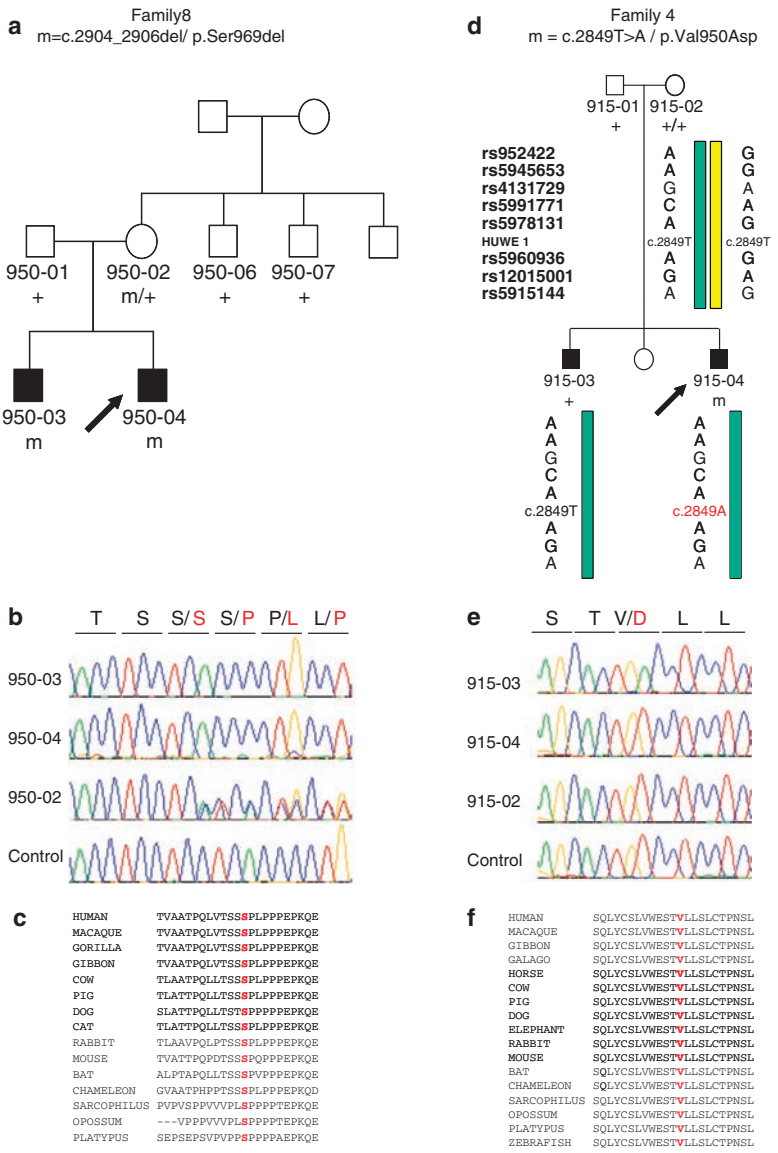


Figure 2 Identification of variants in *PHF8* and *HUWE1* in families 8 and 4. (**a**) Pedigree of family 8 and segregation analysis of the p.Ser969del variant in *PHF8*. The arrow indicates the index case. (**b**) Sequence electropherograms showing the presence of the p.Ser969del variant at the hemizygous state in the two affected brothers and at the heterozygous state in their mother. (**c**) Alignment of the region flanking the variant in orthologous proteins, showing the high conservation of Serine 969. (**d**) Pedigree of family 4 and haplotypes reconstructed from eight informative single nucleotide polymorphisms (SNPs) adjacent to *HUWE1* (genotypes of these SNPs were obtained from Illumina cytoSNP-12 arrays analysis), showing that the same maternal haplotype was transmitted to the affected brothers with and without the p.Val950Asp mutation. The arrow indicates the index case. (**e**) Sequence electropherograms showing the presence of p.Val950Asp in the index case and its absence in the affected brother and in the mother. These results are consistent with the *de novo* occurrence of p.Val950Asp in the index case. (**f**) Alignment of the region flanking the variant in orthologous proteins, showing the high conservation of valine 950.

patients and their heterozygous mother, one with a premature termination codon in exon 3 and one missing exon 3, which restored the reading frame (Figure 3c). Exon 3 skipping could have resulted from nonsense-associated altered splicing, a mechanism alternative to nonsense-mediated decay.[27]

To confirm that mutations in *TMLHE* are associated with ASD, we screened its coding sequence in 501 male probands with ASD. Two missense substitutions (c.730G>C/p.Asp244His and c.1107G>T/p.Glu369Asp) were found in two additional unrelated patients with ASD but not in 330 controls and not reported in databases. Aspartic acid 244 and glutamic acid 369 are highly conserved in other species (Figure 3d). Aspartic acid 244 is predicted to bind the 2-oxoglutarate cofactor and constitutes one of the three key residues of the catalytic core,[28] suggesting

a complete loss-of-function of the protein with this substitution.

To further analyze the consequences of the *TMLHE* mutations, we assayed carnitine and TML, the precursor of carnitine biosynthesis, in plasma and urine from the brothers with the p.Arg77X mutation and the patient with the p.Asp244His variant for whom biological samples were available. Carnitine was slightly but not significantly decreased in the plasma of the patients; the values remained in the normal range (Figure 3f). By contrast, mass spectrometry revealed a significant 2–3-fold increase in the TML precursor in the plasma of all the three patients with the Arg77X and Asp244His mutations compared with controls (Figure 3g). Carnitine intake, estimated from a 3-day dietary recall questionnaire, was normal in all patients (estimated at 55, 80 and 96 mg j^{-1}, respectively).

Very few variants in *TMLHE* predicted to have deleterious effects are present in genetic databases (HapMap, 1000 Genomes); in particular, no nonsense mutations have been identified and only 18 non-synonymous variants have been reported on >8700 X chromosomes in the Exome Variant Server. Furthermore, non-synonymous variants are far more frequent in females than in males ($n=22/3381$ versus $n=3/1998$, if we exclude the Asn235Thr variant specific of the African population), indicating that variations in *TMLHE* are not well tolerated and are liable to be pathogenic. Mutations in *TMLHE* were found in patients with autism and ID but not in patients with Asperger syndrome. If we exclude patients with Asperger syndrome, the difference between point mutations in *TMLHE* in male patients with autism and mental retardation and male individuals from Exome variant server is significant ($P=0.05$).

Interestingly, Celestino-Soper et al.[29] have recently reported that the deletion of exon 2 of *TMLHE* is a CNV present in $\sim$1/350 males and associated with ASD with a low penetrance. To compare the frequency of this CNV in healthy individuals and patients with ASD, we specifically assessed its presence in 896 unrelated healthy male controls and 691 patients with ASD by quantitative multiplex assay or Illumina Human 1M-single BeadChip arrays (see Material and methods). This study revealed the presence of one deletion of exon 2 of *TMLHE* in a single subject out of 896 male controls, whereas it was present in 3 out of 691 male patients with ASD ($P=0.3$). Interestingly, one of the three patients with the deletion of exon 2 had an affected brother who did not carry the deletion. In addition, a duplication encompassing exon 1 of *TMLHE*, absent from 525 healthy controls, was detected by Illumina Human 1M-single BeadChip arrays in an additional patient with ASD ($P=0.4$). Altogether, these results suggest that deficits in *TMLHE* could be rarer than previously reported. These deficits seem to be more frequent in patients with ASD, suggesting that they constitute susceptibility factors for autism, although the difference did not reach significance for microrearrangements in *TMLHE*.

Discussion

This study focused on rare variants on chromosome X in multiplex families with ASD compatible with X-linked inheritance. Altogether, this study identified 36 possibly deleterious variants in 33 genes, including *PHF8*, *HUWE1* and *TMLHE*. Variants in genes common to at least two families were exceptional, confirming that ASD is genetically highly heterogeneous.

The X chromosome contains the largest number of genes expressed in the brain.[30] For this reason, mutations causing monogenic forms of ID have been identified in numerous genes on chromosome X.[31] Interestingly, almost all genes involved in ASD, such as *NLGN3/4X* on chromosome X or *SHANK3* on chromosome 22qter, are also mutated in patients with ID without autistic features.[11,31] In this study, variants in two genes were previously implicated in ID: *PHF8*, which encodes a histone lysine demethylase that regulates rRNA synthesis[32] and retinoic acid-induced neuronal differentiation,[33] and *HUWE1*, which encodes an E3 ubiquitin-protein ligase that controls neural differentiation and proliferation by catalyzing the polyubiquitination and degradation of the *N*-Myc oncoprotein.[34,35] Missense mutations and microduplications encompassing *HUWE1* were identified in a few large families with moderate-to-severe ID,[24] whereas nonsense, truncating and one missense mutations in *PHF8* were previously reported in patients with ID and cleft lip/palate.[21–23] Interestingly, in a previous report, two brothers with a deletion encompassing *PHF8* and two nearby genes (*FAM120C* and *WNK3*) also had autistic features.[36] In our study, both variants alter highly conserved amino acids in the proteins and are predicted to be deleterious; the p.Ser969del variant in *PHF8* segregated in the two affected brothers from family 8 and the p.Val950Asp in *HUWE1* occurred *de novo* in the index case of family 4 who is more severely affected than his brother. The p.Glu441Lys in *PPP1R3F* was also present in the two affected brothers of family 4 but was inherited from their unaffected maternal grandfather. Interestingly, another missense variant (p.Phe245Leu) in *PPP1R3F* was reported in a patient with Asperger syndrome.[17] This observation suggests that the variants identified in family 4 act as risk factors for ASD when associated with other deleterious variants on the X chromosome or autosomes. Further studies are needed to confirm the roles of these genes in ASD.

Among the variants detected in this study, there was a single nonsense mutation in *TMLHE* that segregated with ASD in the two affected brothers of family 9. *TMLHE* is located at the far end of the long arm of chromosome X (Xq28) and encodes ε-*N*-trimethyllysine hydroxylase, the enzyme that catalyzes the first of the four steps of endogenous carnitine biosynthesis. Screening of additional male patients with ASD identified two missense variants predicted to be deleterious in unrelated sporadic patients. The p.Arg77X and p.Asp244His mutations were associated with a significant increase of TML, the substrate of *TMLHE* and precursor of carnitine biosynthesis, in the plasma of the patients, confirming that they lead to loss of *TMLHE* function and deficit of endogenous carnitine biosynthesis. However, although carnitine was mildly decreased, it remained in the normal range in the plasma and urine of the patients. This result is not surprising as, in humans, the carnitine pool mainly comes from food intake. A small pool ($\sim$25%) of carnitine is also synthesized in liver, kidney and brain, but the precise role of endogenous

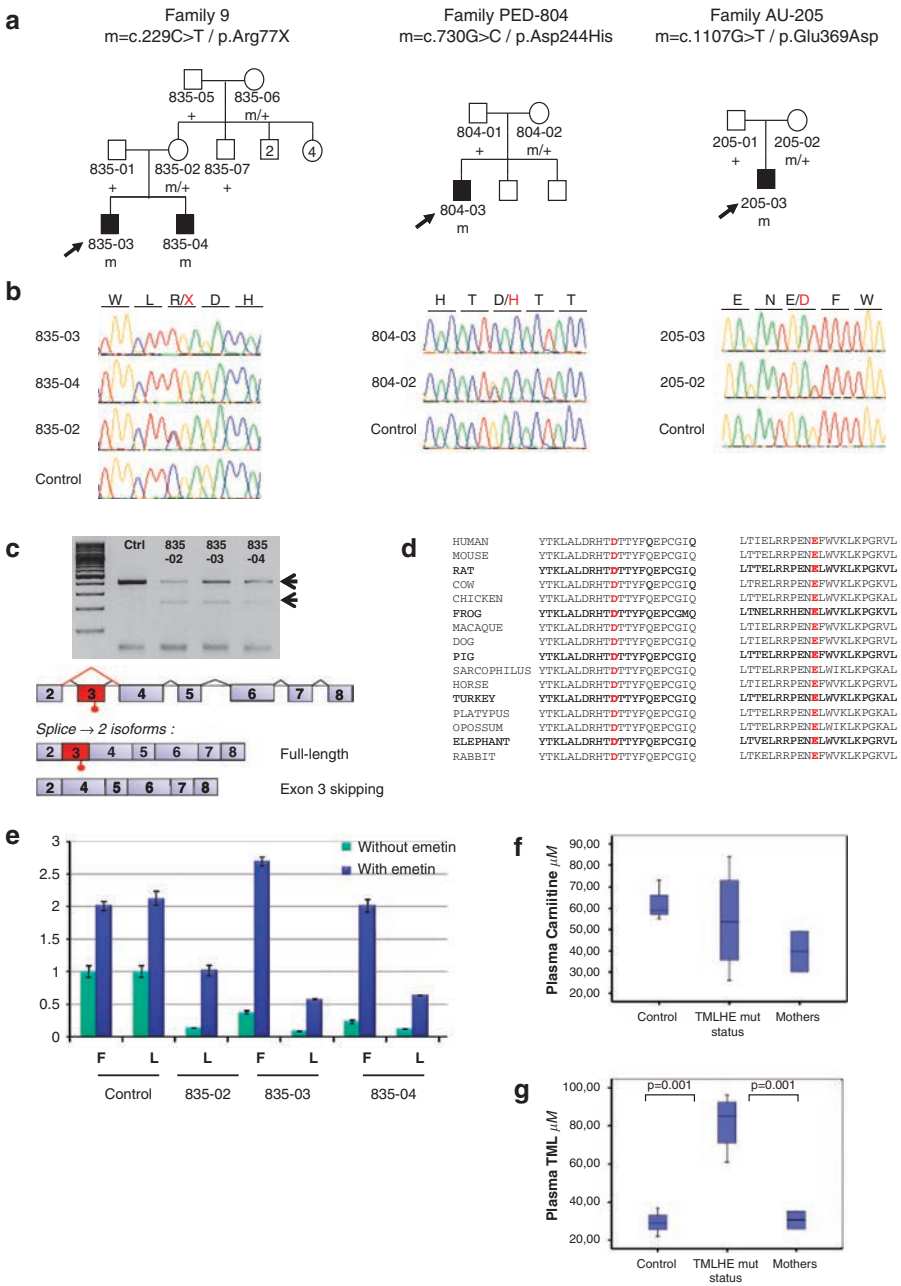


Figure 3 Identification of *TMLHE* mutations in three families. (**a**) Pedigrees and segregation analysis of the *TMLHE* mutations in families 9, PED-804 and AU-205. The arrows indicate the index cases. (**b**) Sequence electropherograms of the mutations at the hemizygous state in the index cases (835–03 in family 9, 804–03 and 205–03) and the affected brother of family 9 (835–04), and at the heterozygous state in the mothers (835–02, 804–02 and 205–02). (**c**) Analysis of *TMLHE* mRNA in lymphoblasts from members of family 9 and schematic representations of the splicing isoforms detected in subjects with the p.Arg77X mutation in exon 3. Reverse transcriptase–PCR products using primer pairs in exons 2 and 4, run on 2% agarose gels, showed two mRNA isoforms in the index case (835–03), his affected brother (835–04) and his mother (835–02) and a single isoform in a control subject (**c**). Sequence analysis confirmed that the long isoform contains the premature termination codon in exon 3 and that exon 3 was skipped in the short isoform, probably as a consequence of nonsense-associated alternative splicing. (**d**) Alignment of the region flanking the two missense variants in orthologous proteins showing the conservation of the altered amino acids. (**e**) Quantification of *TMLHE* mRNA expression in fibroblasts (F) and lymphoblasts (L) from members of family 9 by quantitative real-time PCR, using primer pairs in exons 7 and 8. *TMLHE* mRNA was expressed 10 times less in patients compared with healthy controls (green bars). Overnight treatment with 10 $\mu g\,ml^{-1}$ emetin (blue bars), an inhibitor of nonsense-mediated decay, restored the expression of the *TMLHE* mRNA. (**f**) Assay of free carnitine by UPLC (ultra performance liquid chromatography) chromatographic and TQD (tandem quadrupole detector) mass spectrometry in the plasma of patients. (**g**) Assay of trimethyllysine (TML) by UPLC chromatographic and TQD mass spectrometry showing a 2–3-fold increase in the plasma of patients.

synthesis in these tissues remains unknown.[28] Carnitine is an essential metabolite in all animal species as well as in numerous microorganisms and plants. In mammalian cells, carnitine is present as free carnitine and acylcarnitines, including acetylcarnitine.[37,38] The main role of carnitine is to transport activated long-chain fatty acids across the inner mitochondrial membrane for β-oxidation. But additional neuroprotective, neuromodulatory and neurotrophic roles have also been suggested.[39–42] In particular, carnitine is an antioxidant and that might protect mitochondria from oxidative stress.

Another recent study supports the observation that the loss-of-function of *TMLHE* is associated with ASD. The deletion of exon 2 of *TMLHE*, originally identified in ASD male patients,[43] was shown to be a CNV that is present in male controls at a frequency of ~1/350. The first coding exon of *TMLHE*, exon 2, encodes the signal peptide necessary to address the protein to mitochondria. Deletion of this exon causes enzyme deficiency and impairs endogenous carnitine biosynthesis, as observed for p.Arg77X and p.Asp244His mutations. Remarkably, 6 out of 7 sib pairs with ASD were concordant for the deletion of exon 2, whereas the frequency of this CNV was only slightly, not significantly, increased in sporadic ASD patient. The authors concluded that the deletion of exon 2 of *TMLHE* is a risk factor for nondysmorphic ASD, with a penetrance estimated at 2–4%.[29] In this study, the deletion of exon 2 of *TMLHE* was present in three male patients with ASD out of 691 and only in one male control out of 896, indicating that this CNV could even be rarer than ~1/350. Although CNVs altering *TMLHE* tend to be more frequent in patients with ASD than in controls, the difference of frequencies between patients and controls was not significant. Yet, this finding is concordant with the results of the previous study in which significance between patients and controls was not achieved either.[29] Interestingly, in our case, the only sib pair with ASD was discordant for the deletion of exon 2 of *TMLHE* and all the other patients were sporadic cases. However, association studies are not appropriate for rare variants, especially in the case of a high genetic heterogeneity such as in ASD. The arguments suggesting that a loss-of-function of *TMLHE* could contribute to autism are: (i) the demonstration that mutations have functional biological impact; (ii) the segregation of the mutations with the disorder in the families; and (iii) the identification of different rare functional mutations in the same gene in unrelated individuals.[44] Taking these arguments into account, our results confirm that the deficiency of *TMLHE* likely contributes to ASD, probably in association with other genetic or non-genetic abnormalities, and reinforces the view that point mutations could also be identified in patients with ASD. The penetrance of the point mutations remains to be determined but might be higher than previously anticipated for deletion of exon 2.

The mechanism by which *TMLHE* deficiency leads to ASD remains unclear. The increase in TML could be toxic at some stage during brain development or interfere with the establishment of normal neuronal networks. Alternatively, the deficiency of carnitine itself or one of the three intermediates of endogenous carnitine biosynthesis (HTML, TMABA or γ-BB, Supplementary Figure S4) might be deleterious for brain development, alone or combined with a deficit in carnitine intake. To test the hypothesis that patients with p.Arg77X and p.Asp244His had a deficiency of carnitine intake in addition of the *TMLHE* mutation, we assessed their dietary intake over several days. The calculated carnitine intake was normal in all the three patients but reflects only current carnitine intake; a lack of carnitine during specific antenatal or neonatal periods cannot be ruled out. A retrospective questionnaire revealed that patient PED-804–03 (with the p.Asp244His mutation) refused to eat meat around 12 months of age. However, young children frequently refuse meat, particularly those with ASD. Interestingly, low levels of carnitine and acetylcarnitines and altered brain fatty acid metabolism were reported in subjects with ASD.[45–47] However, if a systemic carnitine deficiency constitutes a risk factor for ASD, infants with low meat intake would have a higher risk of developing ASD. This is unlikely as the prevalence of autism is not notably increased in vegetarians or populations with low meat intake. In addition, carnitine and related metabolites are often abnormal in patients without ASD. These arguments suggest that low levels of carnitine and acetylcarnitines in ASD patients could reflect a more complex deficit. Further studies are, therefore, needed to decipher the real role of carnitine and *TMLHE* deficiency in ASD.

Recent studies have emphasized the role of carnitine in promoting social interactions in animal models. Desert (*Schistocerca gregaria*) and migratory locusts (*Locusta Migratoria*) reversibly change between two phenotypes (solitarious and gregarious) that differ in bodily appearance, physiology, brain size and organization, and behavior. At low population density, locusts in the solitarious phase avoid their congeners; when the population increases, locusts become gregarious and aggregate in migratory swarms. The genomes of the two forms of locust are equivalent. The transformation is driven only by epigenetic regulation.[48] Interestingly, carnitine was recently shown to constitute a key regulatory metabolite in the phase transition in the migratory locust.[49] Remarkably, the only other metabolite that is known to regulate phase transition in desert locusts is serotonin.[50] The link between carnitine and serotonin is unclear, but carnitine has been proposed to act as a neuromodulator in the animal central nervous system; in addition, acylcarnitines could promote the biosynthesis and release of neurotransmitters, including dopamine and melatonin.[51] Although these results were obtained in species very distant from humans, they support the hypothesis that carnitine has a conserved role in socialization during evolution and offers potentially novel insights into the complex role of carnitine and its derivatives acetylcarnitine and acylcarnitines in the brain.

Finally, *TMLHE* could also have other yet unknown functions. An isoform has been reported in which exon 2 is spliced out and an alternative start codon in exon 3 is used.[42] This isoform lacks the mitochondrial targeting signal and probably localizes in another cellular compartment. Contrary to the deletion of exon 2, which affects only the mitochondrial isoform,[29] the mutations reported in this study are predicted to affect other isoforms as well. Interestingly, the TMLHE protein was reported to interact with nuclear complex p130/RBL2 that regulates gene expression, supporting the hypothesis that TMLHE has another cellular localization and function.[52] The mechanism by which the *TMLHE* deficiency leads to ASD remains to be further characterized.

Altogether, our results confirm that a *TMLHE* deficiency is associated with ASD and support the hypothesis that rare variants on the X chromosome are involved in the etiology of ASD and contribute to the sex-ratio disequilibrium characteristic of these disorders.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements. We thank the patients for their participation in the study, the P3S platform and the genotyping and sequencing platform of the ICM for technical assistance and the DNA and cell bank of CRICM for DNA extraction and cell culture. We also thank Dr Merle Ruberg for critical reading of the manuscript and Pr André Mégarbané for kindly providing DNA of controls individuals from Lebanon. This study was financially supported by GIS-Maladies Rares, Fondation de France, ERA-NET NEURON EUHFAUTISM, INSERM and AP-HP.

1. Freitag CM, Staal W, Klauck SM, Duketis E, Waltes R. Genetics of autistic disorders: review and clinical implications. *Eur Child Adolesc Psychiatry* 2010; **19**: 169–178.
2. Lichtenstein P, Carlstrom E, Rastam M, Gillberg C, Anckarsater H. The genetics of autism spectrum disorders and related neuropsychiatric disorders in childhood. *Am J Psychiatry* 2010; **167**: 1357–1363.
3. Hallmayer J, Cleveland S, Torres A, Phillips J, Cohen B, Torigoe T *et al.* Genetic heritability and shared environmental factors among twin pairs with autism. *Arch Gen Psychiatry* 2011; **68**: 1095–1102.
4. Sanders SJ, Ercan-Sencicek AG, Hus V, Luo R, Murtha MT, Moreno-De-Luca D *et al.* Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron* 2011; **70**: 863–885.
5. Depienne C, Moreno-De-Luca D, Heron D, Bouteiller D, Gennetier A, Delorme R *et al.* Screening for genomic rearrangements and methylation abnormalities of the 15q11-q13 region in autism spectrum disorders. *Biol Psychiatry* 2009; **66**: 349–359.
6. Bucan M, Abrahams BS, Wang K, Glessner JT, Herman EI, Sonnenblick LI *et al.* Genome-wide analyses of exonic copy number variants in a family-based study point to novel autism susceptibility genes. *PLoS Genet* 2009; **5**: e1000536.
7. Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R *et al.* Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* 2010; **466**: 368–372.
8. Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A *et al.* Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* 2012; **485**: 242–245.
9. O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP *et al.* Sporadic autism exomes reveal a highly interconnected protein network of *de novo* mutations. *Nature* 2012; **485**: 246–250.
10. Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ *et al.* De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* 2012; **485**: 237–241.
11. Bourgeron T. A synaptic trek to autism. *Curr Opin Neurobiol* 2009; **19**: 231–234.
12. Sato D, Lionel AC, Leblond CS, Prasad A, Pinto D, Walker S *et al.* SHANK1 deletions in males with autism spectrum disorder. *Am J Human Genet* 2012; **90**: 879–887.
13. Leblond CS, Heinrich J, Delorme R, Proepper C, Betancur C, Huguet G *et al.* Genetic and functional analyses of SHANK2 mutations suggest a multiple hit model of autism spectrum disorders. *PLoS Genet* 2012; **8**: e1002521.
14. Gillberg C, Cederlund M, Lamberg K, Zeijlon L. Brief report: 'the autism epidemic'. The registered prevalence of autism in a Swedish urban area. *J Autism Dev Disord* 2006; **36**: 429–435.
15. Jamain S, Quach H, Betancur C, Rastam M, Colineaux C, Gillberg IC *et al.* Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. *Nat Genet* 2003; **34**: 27–29.
16. Tarpey PS, Smith R, Pleasance E, Whibley A, Edkins S, Hardy C *et al.* A systematic, large-scale resequencing screen of X-chromosome coding exons in mental retardation. *Nat Genet* 2009; **41**: 535–543.
17. Piton A, Gauthier J, Hamdan FF, Lafreniere RG, Yang Y, Henrion E *et al.* Systematic resequencing of X-chromosome synaptic genes in autism spectrum disorder and schizophrenia. *Mol Psychiatry* 2011; **16**: 867–880.
18. Ozonoff S, Young GS, Carter A, Messinger D, Yirmiya N, Zwaigenbaum L *et al.* Recurrence risk for autism spectrum disorders: a Baby Siblings Research Consortium study. *Pediatrics* 2011; **128**: e488–495.
19. Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W *et al.* Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol* 2009; **27**: 182–189.
20. Colella S, Yau C, Taylor JM, Mirza G, Butler H, Clouston P, Bassett AS, Seller A, Holmes CC, Ragoussis J. QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. *Nucleic Acids Res* 2007; **35**: 2013–2025.
21. Laumonnier F, Holbert S, Ronce N, Faravelli F, Lenzner S, Schwartz CE *et al.* Mutations in PHF8 are associated with X linked mental retardation and cleft lip/cleft palate. *J Med Genet* 2005; **42**: 780–786.
22. Abidi FE, Miano MG, Murray JC, Schwartz CE. A novel mutation in the PHF8 gene is associated with X-linked mental retardation with cleft lip/cleft palate. *Clin Genet* 2007; **72**: 19–22.
23. Koivisto AM, Ala-Mello S, Lemmela S, Komu HA, Rautio J, Jarvela I. Screening of mutations in the PHF8 gene and identification of a novel mutation in a Finnish family with XLMR and cleft lip/cleft palate. *Clin Genet* 2007; **72**: 145–149.
24. Froyen G, Corbett M, Vandewalle J, Jarvela I, Lawrence O, Meldrum C *et al.* Submicroscopic duplications of the hydroxysteroid dehydrogenase HSD17B10 and the E3 ubiquitin ligase HUWE1 are associated with mental retardation. *Am J Human Gene* 2008; **82**: 432–443.
25. Kenzelmann D, Chiquet-Ehrismann R, Leachman NT, Tucker RP. Teneurin-1 is expressed in interconnected regions of the developing brain and is processed *in vivo*. *BMC Dev Biol* 2008; **8**: 30.
26. Pilotte J, Dupont-Versteegden EE, Vanderklish PW. Widespread regulation of miRNA biogenesis at the Dicer step by the cold-inducible RNA-binding protein, RBM3. *PLoS ONE* 2011; **6**: e28446.
27. Cartegni L, Chew SL, Krainer AR. Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev* 2002; **3**: 285–298.
28. Monfregola J, Cevenini A, Terracciano A, van Vlies N, Arbucci S, Wanders RJ *et al.* Functional analysis of TMLH variants and definition of domains required for catalytic activity and mitochondrial targeting. *J Cell Physiol* 2005; **204**: 839–847.
29. Celestino-Soper PB, Violante S, Crawford EL, Luo R, Lionel AC, Delaby E *et al.* A common X-linked inborn error of carnitine biosynthesis may be a risk factor for nondysmorphic autism. *Proc Natl Acad Sci USA* 2012; **109**: 7974–7981.
30. Nguyen DK, Disteche CM. High expression of the mammalian X chromosome in brain. *Brain Res* 2006; **1126**: 46–49.
31. Raymond FL. X linked mental retardation: a clinical guide. *J Med Genet* 2006; **43**: 193–200.
32. Zhu Z, Wang Y, Li X, Wang Y, Xu L, Wang X *et al.* PHF8 is a histone H3K9me2 demethylase regulating rRNA synthesis. *Cell Res* 2010; **20**: 794–801.
33. Qiu J, Shi G, Jia Y, Li J, Wu M, Li J *et al.* The X-linked mental retardation gene PHF8 is a histone demethylase involved in neuronal differentiation. *Cell Res* 2010; **20**: 908–918.
34. Zhao X, Heng JI, Guardavaccaro D, Jiang R, Pagano M, Guillemot F *et al.* The HECT-domain ubiquitin ligase Huwe1 controls neural differentiation and proliferation by destabilizing the N-Myc oncoprotein. *Nat Cell biol* 2008; **10**: 643–653.
35. D'Arca D, Zhao X, Xu W, Ramirez-Martinez NC, Iavarone A, Lasorella A. Huwe1 ubiquitin ligase is essential to synchronize neuronal and glial differentiation in the developing cerebellum. *Proc Natl Acad Sci USA* 2010; **107**: 5875–5880.
36. Qiao Y, Liu X, Harvard C, Hildebrand MJ, Rajcan-Separovic E, Holden JJ *et al.* Autism-associated familial microdeletion of Xp11.22. *Clin Genet* 2008; **74**: 134–144.
37. Vaz FM, Wanders RJ. Carnitine biosynthesis in mammals. *Biochem J* 2002; **361**(Pt 3): 417–429.
38. Jones LL, McDonald DA, Borum PR. Acylcarnitines: role in brain. *Prog Lipid Res* 2010; **49**: 61–75.
39. Virmani A, Binienda Z. Role of carnitine esters in brain neuropathology. *Mol Aspects Med* 2004; **25**: 533–549.
40. Nalecz KA, Miecz D, Berezowski V, Cecchelli R. Carnitine: transport and physiological functions in the brain. *Mol Aspects Med* 2004; **25**: 551–567.
41. Nalecz KA, Nalecz MJ. Carnitine–a known compound, a novel function in neural cells. *Acta Neurobiol Exp* 1996; **56**: 597–609.
42. Monfregola J, Napolitano G, Conte I, Cevenini A, Migliaccio C, D'Urso M *et al.* Functional characterization of the TMLH gene: promoter analysis, in situ hybridization, identification and mapping of alternative splicing variants. *Gene* 2007; **395**: 86–97.
43. Celestino-Soper PB, Shaw CA, Sanders SJ, Li J, Murtha MT, Ercan-Sencicek AG *et al.* Use of array CGH to detect exonic copy number variants throughout the genome in autism families detects a novel deletion in TMLHE. *Human Mol Genet* 2011; **20**: 4360–4370.
44. McClellan J, King MC. Genetic heterogeneity in human disease. *Cell* 2010; **141**: 210–217.
45. Clark-Taylor T, Clark-Taylor BE. Is autism a disorder of fatty acid metabolism? Possible dysfunction of mitochondrial beta-oxidation by long chain acyl-CoA dehydrogenase. *Med Hypotheses* 2004; **62**: 970–975.
46. Filipek PA, Juranek J, Nguyen MT, Cummings C, Gargus JJ. Relative carnitine deficiency in autism. *J Autism Dev Disord* 2004; **34**: 615–623.
47. Rossignol DA, Frye RE. Mitochondrial dysfunction in autism spectrum disorders: a systematic review and meta-analysis. *Mol Psychiatry* 2012; **17**: 290–314.
48. Burrows M, Rogers SM, Ott SR. Epigenetic remodelling of brain, body and behaviour during phase change in locusts. *Neural Syst Circuits* 2011; **1**: 11.

49. Wu R, Wu Z, Wang X, Yang P, Yu D, Zhao C *et al.* Metabolomic analysis reveals that carnitines are key regulatory metabolites in phase transition of the locusts. *Proc Natl Acad Sci USA* 2012; **109**: 3259–3263.
50. Anstey ML, Rogers SM, Ott SR, Burrows M, Simpson SJ. Serotonin mediates behavioral gregarization underlying swarm formation in desert locusts. *Science (New York, NY* 2009; **323**: 627–630.
51. Scafidi S, Fiskum G, Lindauer SL, Bamford P, Shi D, Hopkins I *et al.* Metabolism of acetyl-L-carnitine for energy and neurotransmitter synthesis in the immature rat brain. *J Neurochem* 2010; **114**: 820–831.
52. Litovchick L, Sadasivam S, Florens L, Zhu X, Swanson SK, Velmurugan S *et al.* Evolutionarily conserved multisubunit RBL2/p130 and E2F4 protein complex represses human cell cycle-dependent genes in quiescence. *Mol Cell* 2007; **26**: 539–551.

Supplementary Information accompanies the paper on the Translational Psychiatry website (http://www.nature.com/tp)

SHANK1 Deletions in Males with Autism Spectrum Disorder

Daisuke Sato,[1] Anath C. Lionel,[1,2] Claire S. Leblond,[3,4,5] Aparna Prasad,[1] Dalila Pinto,[1] Susan Walker,[1] Irene O'Connor,[6] Carolyn Russell,[6] Irene E. Drmic,[7] Fadi F. Hamdan,[8] Jacques L. Michaud,[8] Volker Endris,[9] Ralph Roeth,[9] Richard Delorme,[3,4,5,10] Guillaume Huguet,[3,4,5] Marion Leboyer,[11,12] Maria Rastam,[13] Christopher Gillberg,[14,15] Mark Lathrop,[16] Dimitri J. Stavropoulos,[17] Evdokia Anagnostou,[18] Rosanna Weksberg,[19] Eric Fombonne,[20] Lonnie Zwaigenbaum,[21] Bridget A. Fernandez,[22] Wendy Roberts,[7,18] Gudrun A. Rappold,[9] Christian R. Marshall,[1,2] Thomas Bourgeron,[3,4,5] Peter Szatmari,[6,*] and Stephen W. Scherer[1,2,*]

Recent studies have highlighted the involvement of rare (<1% frequency) copy-number variations and point mutations in the genetic etiology of autism spectrum disorder (ASD); these variants particularly affect genes involved in the neuronal synaptic complex. The SHANK gene family consists of three members (*SHANK1*, *SHANK2*, and *SHANK3*), which encode scaffolding proteins required for the proper formation and function of neuronal synapses. Although *SHANK2* and *SHANK3* mutations have been implicated in ASD and intellectual disability, the involvement of *SHANK1* is unknown. Here, we assess microarray data from 1,158 Canadian and 456 European individuals with ASD to discover microdeletions at the *SHANK1* locus on chromosome 19. We identify a hemizygous *SHANK1* deletion that segregates in a four-generation family in which male carriers—but not female carriers—have ASD with higher functioning. A de novo *SHANK1* deletion was also detected in an unrelated male individual with ASD with higher functioning, and no equivalent *SHANK1* mutations were found in >15,000 controls ($p = 0.009$). The discovery of apparent reduced penetrance of ASD in females bearing inherited autosomal *SHANK1* deletions provides a possible contributory model for the male gender bias in autism. The data are also informative for clinical-genetics interpretations of both inherited and sporadic forms of ASD involving *SHANK1*.

Autism is the prototypic form of a group of conditions, also known as "autism spectrum disorders" (ASDs [MIM 209850]), that share common characteristics (impairments in socialization and communication and a pattern of repetitive interests and behaviors) but differ in developmental course, symptom pattern, and cognitive and language abilities. Other ASD subtypes include Asperger disorder (which has less severe language and cognitive deficits) and pervasive developmental disorder not otherwise specified (PDD-NOS; subthreshold symptoms and/or later onset). Subclinical forms of ASD are often characterized as the broader autism phenotype (BAP).[1] Twin and family studies provide evidence of the importance of complex genetic factors in the development of both sporadic and inherited forms of idiopathic autism. An enigma in ASD is the 4:1 male to female gender bias, which might rise to 11:1 when Asperger disorder is considered.[2]

Rare copy-number variations (CNVs) and sequence-level mutations have been identified as etiologic factors in ASD.[3–5] De novo CNVs are observed in 5%–10% of ASD cases.[6–10] A relative enrichment of CNVs disrupting synaptic complex genes[11,12] is observed, and *NLGN3* (MIM 300336), *NLGN4* (MIM 300427),[13] *NRXN1* (MIM 600565),[14] *NRXN3* (MIM 600567),[15] *SHANK2* (MIM 603290),[9,16,17] and *SHANK3* (MIM 606230)[18–20] have been identified as highly penetrant susceptibility loci for ASD and intellectual disability (ID). The SHANK gene family,[21,22] which also includes *SHANK1* (MIM 604999), encodes scaffolding proteins that localize to postsynaptic sites of excitatory synapses in the brain.

We describe seven individuals with deletions involving *SHANK1*. Four male cases have ASD with higher functioning or the BAP and are from a multigenerational family (see family 1 in Figure 1) that carries inherited gene deletions,

[1]The Centre for Applied Genomics and Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, ON M5G 1L7, Canada; [2]Department of Molecular Genetics and McLaughlin Centre, University of Toronto, Toronto, ON M5G 1L7, Canada; [3]Human Genetics and Cognitive Functions, Institut Pasteur, 75015 Paris, France; [4]University Paris Diderot, Sorbonne Paris Cité, Human Genetics and Cognitive Functions, 75013 Paris, France; [5]URA 2182 "Genes, Synapses, and Cognition," Centre National de la Recherche Scientifique, Institut Pasteur, 75015 Paris, France; [6]Offord Centre for Child Studies, Department of Psychiatry and Behavioural Neurosciences, McMaster University, Hamilton, ON L8S 4K1, Canada; [7]Autism Research Unit, The Hospital for Sick Children, Toronto, ON M5G 1X8, Canada; [8]Centre of Excellence in Neuroscience of the Université de Montréal, Sainte-Justine Hospital Research Centre, Montreal, QC H2L 2W5, Canada; [9]Department of Human Molecular Genetics, Heidelberg University, 69120 Heidelberg, Germany; [10]Department of Child and Adolescent Psychiatry, Robert Debré Hospital, Assistance Publique-Hôpitaux de Paris, 75019 Paris, France; [11]Psychiatry Genetic Team, Institut National de la Santé et de la Recherche Médicale U955, 94010 Creteil, France; [12]Fondation FondaMental, 94000, Creteil, France; [13]Department of Clinical Sciences in Lund, Lund University, SE-221 00 Lund, Sweden; [14]Gillberg Neuropsychiatry Centre, University of Gothenburg, SE- 411 19 Göteborg, Sweden; [15]Institute of Child Health, University College London, London WC1N 1EH, UK; [16]Centre National de Génotypage, 91057 Evry, France; [17]Department of Pediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, ON M5G 1L7. Canada; [18]Bloorview Research Institute, University of Toronto, Toronto, ON M4G 1R8, Canada; [19]Program in Genetics and Genome Biology and Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, Toronto, ON M5G 1X8, Canada; [20]Department of Psychiatry, Montreal Children's Hospital and McGill University, Montreal, QC H3Z 1P2, Canada; [21]Department of Pediatrics, University of Alberta, Edmonton, AB T5G 0B7, Canada; [22]Disciplines of Genetics and Medicine, Memorial University of Newfoundland, St. John's, NL A1B 3V6, Canada

*Correspondence: stephen.scherer@sickkids.ca (S.W.S.), szatmar@mcmaster.ca (P.S.)

DOI 10.1016/j.ajhg.2012.03.017.

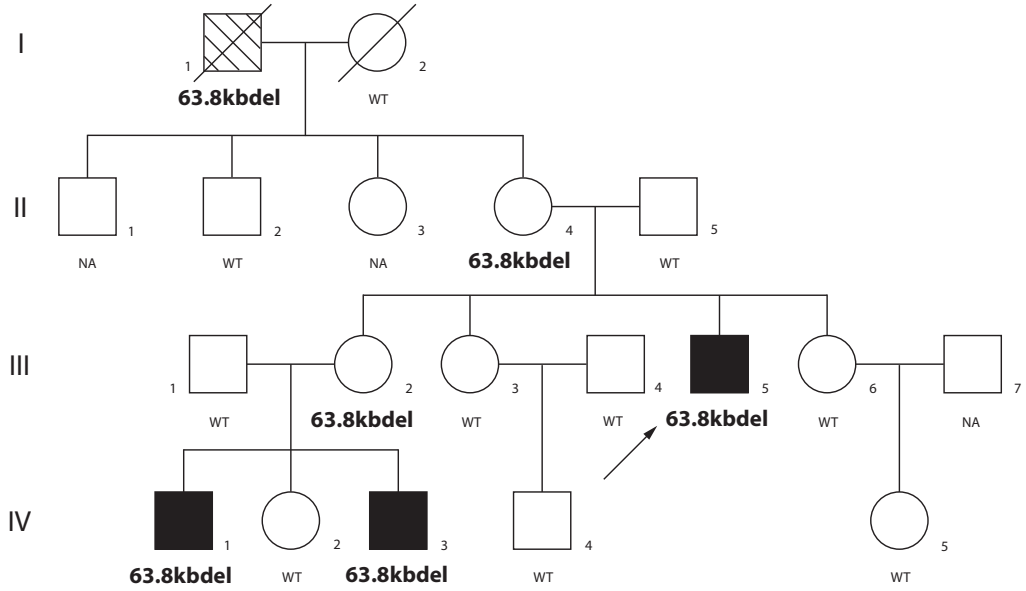


Figure 1. Pedigree of Family 1
Pedigree of a multigenerational family carrying a rare CNV that deletes one copy of *SHANK1*. Individuals with ASD and BAP are indicated by filled symbols and striped symbols, respectively. The proband is indicated by an arrow. "WT" indicates individuals that have the typical copy number of two at the *SHANK1* locus, and "NA" indicates the unavailability of DNA.

an unrelated fifth ASD male case has a de novo deletion at the same locus (see family 2 in Figure S1, available online), and two female individuals from family 1 (Figure 1) have the *SHANK1* deletion but no ASD or BAP.

The ASD dataset examined in this study was composed of 1,158 unrelated Canadian individuals (898 males and 260 females) and 456 unrelated European individuals (362 males and 94 females). All individuals with ASD were diagnosed by expert clinicians on the basis of the Autism Diagnostic Interview-Revised (ADI-R) and/or the Autism Diagnostic Observation Schedule (ADOS).[23] Canadian cases were recruited from five different sites: The Hospital for Sick Children, Toronto, Ontario; McMaster University, Hamilton, Ontario; Memorial University of Newfoundland, St. John's, Newfoundland; University of Alberta, Edmonton, Alberta; and the Montreal Children's Hospital of the McGill University Health Centre, Montreal, Quebec. The European ASD subjects were recruited by the PARIS (Paris Autism Research International Sibpair) study and several other sites at specialized clinical centers dispersed in France, Sweden, Germany, Finland, and the UK. In Sweden, for some cases, the Diagnostic Interview for Social and Communication Disorders (DISCO-10) was applied instead of the ADI-R. The ID dataset consisted of 185 mostly French Canadians (98 males and 87 females) and 155 German nonsyndromic ID cases (93 males and 62 females). Further descriptions of these datasets and the assessment procedures used are available elsewhere.[16,24] Approval was obtained from the research ethics boards at The Hospital for Sick Children (Toronto) and McMaster University for the study, and informed written consent was obtained from all participants.

For the assessment of the presence of CNVs on a genome-wide scale, DNA from the Canadian ASD dataset was genotyped at The Centre for Applied Genomics, Toronto with one of three high-resolution microarray platforms: Affymetrix GeneChip SNP 6.0, Illumina Infinium 1M single SNP, or Agilent SurePrint G3 Human CGH (comparative genomic hybridization) 1x1M. The probe coverage of these platforms at the *SHANK1* region is shown in Figure S2. CNVs were analyzed with published methods.[9,25,26] Briefly, CNV calling was performed with a multialgorithm approach that used at least two different algorithms for the three array platforms: Affymetrix 6.0 (Birdsuite, iPattern, and Genotyping Console), Illumina 1M (PennCNV, QuantiSNP, and iPattern), and Agilent CGH 1x1M (DNA Analytics and CBS from DNAcopy package). Subsequent analyses focused on those CNVs spanning five or more array probes and detected by at least two algorithms. Independent validation of the deletion at the *SHANK1* locus in family 1 was performed with SYBR-Green-based real-time quantitative PCR (qPCR), for which two independent primer pairs were placed at the *SHANK1* locus and the *FOXP2* (MIM 605317) locus as a negative (diploid) control (Figure S3). All primers used in qPCR validation and breakpoint mapping are listed in Table S1.

Using the Illumina Human 1M-Duo BeadChip, we genotyped DNA from the European ASD dataset at the Centre National de Génotypage at the Institut Pasteur. CNVs were analyzed with previously published methods.[9]

Table 1. Clinical Description of Individuals Carrying *SHANK1* Deletion

	Clinical Details				
	Diagnosis	**IQ**[a]	**Language**[b]	**Adaptive Behavior**[c]	**Brain Imaging**
Family 1[d]					
III-5 (male)	ASD[e]: Asperger disorder (ADI-R and ADOS-4) and anxiety	Leiter-R: brief NVIQ = 83 (13^{th} % ile; LA)	OWLS: TL = 68 (2^{nd} % ile; delay)	VABS-I: ABC = 52 ($<1^{st}$ % ile), COM = 43 ($<1^{st}$ % ile), DLS = 63 (1^{st} % ile), and SOC = 65 (1^{st} % ile); he currently takes olanzapine and paroxetine for the anxiety disorder	–
I-1 (male)	BAP: shy, reserved, and reluctant to approach people; amassed a large stamp collection; deceased	–	–	–	–
IV-1 (male)	ASD: Asperger disorder (ADI-R; SRS: 68*T*/mild to moderate)	WASI: VIQ = 114 (82^{nd} % ile; HA) and PIQ = 86 (18^{th} % ile; LA)	OWLS: TL = 93, RL = 82 (12^{th} % ile), and EL = 107 (68^{th} % ile) PPVT: RV = 97 (42^{nd} % ile)	VABS-II: ABC = 85 (16^{th} % ile), COM = 92 (30^{th} % ile), DLS = 85 (16^{th} % ile), and SOC = 85 (16^{th} % ile)	–
IV-3 (male)	ASD: Asperger disorder (ADI-R and ADOS-3)	WPPSI: FSIQ = 89 (23^{rd} % ile; LA), VIQ = 89 (23^{rd} % ile), and PIQ = 91 (27^{th} % ile)	OWLS: TL = 80 (9^{th} % ile), RL = 78 (7^{th} % ile), and EL = 86 (18^{th} % ile) PPVT: RV = 91 (27^{th} % ile)	VABS-II: ABC = 86 (18^{th} % ile), COM = 91 (27^{th} % ile), DLS = 89 (23^{rd} % ile), SOC = 85 (16^{th} % ile), and MOT = 91 (27^{th} % ile)	–
II-4 (female)	non-ASD and non-BAP; anxiety and shyness	–	–	–	–
III-2 (female)	non-ASD and non-BAP; social anxiety disorder and generalized anxiety disorder; shy as a child	–	PPVT: RV = 111 (77^{th} % ile)	–	–
Family 2[f]					
II-1 (male)	ASD with higher functioning (ADI-R; CARS: mild autism)	WISC: FSIQ = 115 (84^{th} % ile; HA), VIQ = 120 (93^{rd} % ile), and PIQ = 100 (50^{th} % ile) (VIQ > PIQ)	–	–	PET: mild hyperfusion temporal left

Refer to pedigrees in Figure 1 (family 1) and Figure S1 (family 2). The following abbreviations are used: IQ, intelligence quotient; ASD, autism spectrum disorder; % ile, percentile; and PET, positron emission tomography.

[a]IQ was measured with an age-appropriate Weschler scale (WPPSI, Wechsler Preschool and Primary Scale of Intelligence; WISC, Wechsler Intelligence Scale for Children; or WASI, Wechsler Abbreviated Scale of Intelligence). Standard scores and percentiles are presented for full-scale IQ (FSIQ), verbal IQ (VIQ), and/or performance IQ (PIQ). FSIQ is not a valid estimate of IQ when significant discrepancy exists between VIQ and PIQ. Leiter International Performance Scale-Revised (Leiter-R) is a measure of nonverbal IQ (NVIQ) only. Percentile classifications are the following: very superior (VS; $>98^{th}$ % ile), superior (S; 91^{st}–97^{th} % ile), high average (HA; 75^{th}–90^{th} % ile), average (A; 25^{th}–74^{th} % ile), low average (LA; 9^{th}–24^{th} % ile), borderline (B; 2^{nd}–8^{th} % ile), and extremely low (EL; $<2^{nd}$ % ile).

[b]Language was measured with the Oral and Written Language Scales (OWLS). Standard scores and percentiles are presented for total language (TL), receptive language (RL), and/or expressive language (EL). Language was rated as nonverbal, average, or delayed ($\leq 16^{th}$ % ile). The Peabody Picture Vocabulary Test (PPVT-4^{th} edition) measured receptive vocabulary (RV).

[c]Adaptive Behavior was measured with the Vineland Adaptive Behavior Scales (VABS). Standard score and percentiles are presented for adaptive behavior composite (ABC), communication (COM), daily living skills (DLS), socialization (SOC), and motor (MOT; only for children 7 years old or younger).

[d]Language details for IV-2 (female) in family 1 are the following: (OWLS) RL = 87 (19^{th} % ile) and EL = 108 (70^{th} % ile) and (PPVT) RV = 103 (58^{th} % ile).

[e]The autism-spectrum diagnosis is based on the Autism Diagnostic Interview-Revised (ADI-R) and the Autism Diagnostic Observation Schedule (ADOS; one of four possible modules is administered on the basis of age and language level). In some cases, the Social Responsiveness Scale (SRS) was administered, and reported T-scores represent average skills ($\leq$59*T*), mild to moderate concerns (60*T*–75*T*), or a severe range (76*T* or higher). Also, the diagnosis for II-1 in family 2 was based on the Childhood Autism Rating Scale (CARS).

[f]II-3 (female) in family 2 was diagnosed with ASD (ADI-R; CARS: mild autism). Her WIPPSI IQ details are the following: PIQ = 99 (50^{th} % ile) and VIQ = nonfunctional language.

Validation of the array CNV calls was performed with qPCR in a similar way as described above; two independent primer pairs were placed at the *SHANK1* locus and at the exon 18 locus of *SHANK1* as a negative (diploid) control (Figure S3).

We initially identified a hemizygous microdeletion in chromosomal region 19q13.33 in ASD proband III-5 from the Canadian cohort (see family 1 in Table 1 and Figure 1). The deletion, which was determined to be 63.8 kb, eliminated exons 1–20 of *SHANK1* and the

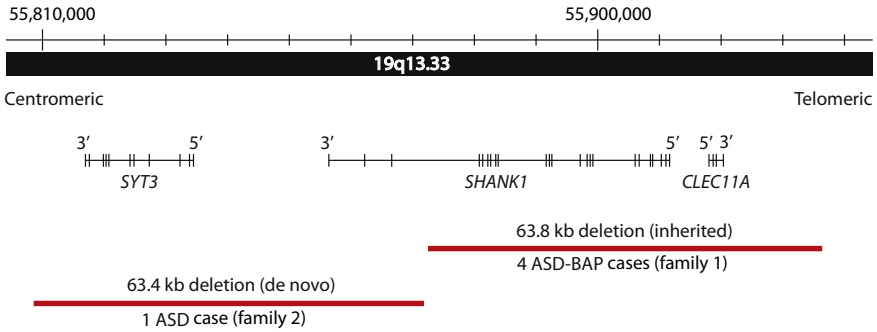


Figure 2. Rare Deletions at *SHANK1* Locus in the Two ASD-Affected Families
Chromosomal position of rare deletions of *SHANK1* and adjacent genes in ASD. The accurate coordinates for family 1 were mapped by sequencing across the breakpoints and are Chr 19: 55,872,189–55,935,995 (hg18). The de novo deletion of family 2 was detected by microarray and has coordinates Chr 19: 55,808,307–55,871,709 (hg18).

neighboring *CLEC11A* (MIM 604713), which codes for a growth factor for primitive hematopoietic progenitor cells (Figure 2). Subsequent genotyping in family 1 revealed that the deletion was also present in males I-1, IV-1, and IV-3 as well as in females II-4 and III-2 (Figure 1).

At 16 years of age, proband III-5 was first assessed by a child psychiatrist and was initially given a clinical diagnosis of PDD-NOS. Starting at an early age, there was evidence of impairment in social communication, but there were not enough repetitive stereotyped behaviors for a diagnosis of autism or Asperger disorder. The ADI-R and the ADOS were completed when the proband was 25 years old. The ADI-R indicated that the parents first became concerned when their son was 12–24 months old, a period during which III-5 engaged in repetitive play and speech. He spoke in single words at 24 months of age and spoke in phrases by 36 months of age. He has never lost language or other skills. He has no history of echolalia, pronoun reversal, or neologisms. His eye contact has always been poor, and he has persistently lacked social smiling, facial affect, joint attention, and empathy. His interests during childhood and adolescence included video games, movies, and sports cards. He graduated from high school, and now, at age 32, he lives independently and works in a sheltered workshop. When III-5 was 25 years old, the ADI-R and ADOS diagnosed him with autism according to the cut-offs, but his current best-estimate diagnoses are that of Asperger disorder (in view of the normal language development) and a separate anxiety disorder; anxiety disorders are common comorbid conditions but are not considered part of the ASD clinical spectrum.

An extensive battery of questionnaires and tests were administered to III-5's parents, and both scored in the typical range. His mother (II-4) has exhibited anxiety and shyness for most of her life but would not be considered to have ASD or BAP. His 40-year-old sister (III-2) is married with one son (IV-1) with Asperger disorder, a neurotypical daughter (IV-2), and a son (IV-3) with ASD. III-2 completed university and worked as a school teacher for years. She has a diagnosis of social anxiety and a generalized anxiety disorder for which she has taken antianxiety medication. Assessment by interview and questionnaire indicated that she was typical for all measures and did not show evidence of ASD or the BAP.

III-5's maternal grandfather (I-1) passed away when he was 95 years old. Little is known about his childhood other than that he had difficulty in school. Throughout his adult life, he had been quiet and withdrawn. He did not develop close relationships. He was a truck driver and spent most of his time away from home. There was no history of psychiatric illness. He was an avid stamp collector but did not use this interest to engage in social interactions. His children were interviewed so that a diagnosis of ASD could be determined; based on this interview, our clinical impression is that he most likely had, at the very least, the BAP or, possibly, high-functioning ASD, which could not be determined under the circumstances.

IV-1 was clinically diagnosed with Asperger disorder when he was 8 years old. He was born by cesarean section 10 days late. Early developmental milestones were within normal limits. When he was 3 years old, his parents detected developmental differences by noting that he was not interested in other children and was preoccupied with objects. He had an encyclopedic knowledge of cars. He would approach other children but tended to play beside them and became upset with changes in routine. He exhibited difficulties with eye contact and understanding social cues and rules.

Additional assessments were conducted when he was 10 years old. He met all the cut-offs for an autistic disorder on the ADI-R except for the nonverbal total. The ADOS scores were below the cut-off for a diagnosis of ASD because of his strengths in the communication domain. Descriptive gestures were present but were vague and infrequent, accounting for his communication score of 1 (the cut-off is 2). Impairments in reciprocal social interaction continued to be evident. His score on the performance intelligence quotient (IQ) component (score = 86)

of psychometric testing was significantly lower than his score on the verbal IQ subtests (score = 114). IV-1 qualified for a diagnosis of Asperger disorder.

Individual IV-3 was first evaluated when he was 3 years old. When he was 18 months old, his parents became concerned because he was not talking. He developed single words at 24 months. He communicated by leading his parents by the hand and exhibited repetitive behaviors. He did not offer comfort or empathy and did not initiate social interaction, although he would play with his parents. Certain noises such as the washing machine or the toilet flushing would bother him; he became upset if his mother had her hair down or a jacket unzipped.

Assessment when he was 5 years and 8 months old indicated that he was positive on the ADI-R for autism and for ASD on the ADOS. He had made good progress in social interaction and language. His expressive language consisted of short sentences and phrases and some echoed speech and mild articulation difficulties. His IQ and expressive and receptive language scores were in the low-average range, leading to a best-estimate diagnosis of Asperger disorder.

In separate microarray experiments examining 456 ASD-affected individuals from Europe, we identified a 63.4 kb hemizygous CNV in individual F2-II-1 from family 2; this CNV deleted the last three exons of *SHANK1* and the entire centromeric synaptotagmin-3 (*SYT3* [MIM 600327]) gene, which plays a role in Ca^{2+}-dependent exocytosis of secretory vesicles (Table 1, Figure 2). Haplotype analysis revealed that the deletion resided on the chromosome originating from the mother (who was shown to carry two copies of *SHANK1*), so it was presumed to be a de novo event (Figure S1). The deletion was not in F2-II-3.

Male individual F2-II-1 (Table 1 and Figure S1) was the first child born to a 20-year-old mother. He has a younger maternal half-sister (F2-II-3) with autism and mild ID. F2-II-1 was born two months before term. Developmental abnormalities were identified during his first year. He did not babble, made no eye contact, and refused to be touched. He started to walk at two years of age, but his motor coordination was poor. He started to talk at 2.5 years of age, which astonished the parents because until then, he had been extremely quiet. He developed a formal, pedantic style of speech with abnormal prosody. He was uninterested in other children. He repeated routines and rituals and accumulated facts on certain subjects such as astronomy. When upset, he flapped his hands or moved his body in a stereotypic fashion. Lately, he has had periods of depression. His IQ is in the normal range and shows that he has good verbal ability. The best estimate diagnosis for F2-II-1 is autism with higher functioning.

No deletion equivalent to those described in families 1 or 2 was observed in 15,122 control individuals. These controls included 2,026 healthy individuals from the Children's Hospital of Philadelphia,[27] 2,493 controls genotyped at the University of Washington,[28] and 10,603 population-based controls[8,9,25] whose microarray data were analyzed by our group. This latter dataset included 1,123 controls from northern Germany,[29] 1,234 Canadian controls from Ottawa,[30] 1,120 population controls from Ontario,[31] 1,056 HapMap samples,[32] 4,783 controls from the Wellcome Trust Case Control Consortium,[33] and 1,287 controls recruited by the Study of Addiction: Genetics and Environment consortium.[34] Control individuals were predominantly of European ancestry, which was comparable to the ancestry of the ASD subjects. We also examined the Database of Genomic Variants[35,36] for previously reported CNVs at the *SHANK1* locus. One study[37] reported an unvalidated 109.1 kb deletion encompassing the entire *SHANK1* gene in a Japanese female HapMap individual (NA18942). We obtained this DNA sample and performed qPCR by using two independent primer pairs at the *SHANK1* locus; we were unable to confirm the deletion, indicating that it was a false-positive call (Figure S3).

The frequency of deletions at the *SHANK1* locus is significantly higher in ASD cases than in controls (2/1,614 cases versus 0/15,122 controls; Fisher's Exact test two-tailed p = 0.009). No other obvious potentially etiologic CNV was observed in any of the individuals with ASD in families 1 or 2 (Tables S2 and S3). Therefore, at this resolution of analysis, the rare deletions of common segments of *SHANK1* were the only common events observed between the two unrelated ASD families.

To test for potential damaging sequence-level mutations in *SHANK1*, we used Sanger sequencing to examine all 23 exons and splice sites in 509 unrelated ASD (384 male and 125 female) and 340 ID (191 males and 149 females) individuals. The rationale for *SHANK1* mutation screening of ASD and ID cases arises from our previous observation of rare truncating nonsense and/or frameshift mutations in *SHANK2*[16] and *SHANK3*[18,19,38] in both of these disorders. We used Primer3 software v.0.4.0 to design PCR primers. We performed PCRs by using standard conditions, and we purified and sequenced products directly by using the BigDye Terminator sequencing (Applied Biosystems, Foster City, CA, USA). Variant detection was performed with SeqScape software from Applied Biosystems. We validated variants detected in the cases but not previously reported in the Single Nucleotide Polymorphism Database (dbSNP) build 130 by resequencing samples from the proband, both parents, and the siblings when available. All primers used for Sanger sequencing of *SHANK1* and *PCDHGA11* are listed in Table S4.

We detected 26 rare missense variants in 23 ASD and seven ID cases, which were not found in dbSNP build 130 or in 285 control individuals from the Ontario general population (Table S5 and Figures S4 and S5). However, only two of these missense variants (c.877 G>A [p.Asp293Asn] in families 5 and 6 and c.2207 G>A [p.Arg736Gln] in family 9) are predicted to be damaging on the basis of their alteration of highly conserved residues within the ANK and PDZ domains, respectively. Although they occur in males with ASD, both variants are also found in non-ASD

(or BAP) fathers. No significant mutation was found on the nondeleted allele of proband III-5.

We also conducted whole-exome sequencing in individuals III-5 and IV-3 from family 1 to search for potential mutations in other genes (Table S6). We performed target enrichment by utilizing the Agilent SureSelect 50Mb Human All Exon kit (Agilent Technologies, Santa Clara, CA, USA), and we conducted paired-end sequencing on a Life Technologies SOLiD5500XL (Life Technologies, Foster City, CA, USA) platform. Protocols for sequencing and target capture followed specifications from the manufacturers.

A nonsense mutation (Tyr313*) predicted to introduce a stop codon in the *PCDHGA11* prodocadherin gene in chromosomal region 5q31.3 was identified. *PCDHGA11* is a member of the protocadherin gamma gene cluster thought to have an important role in establishing connections in the brain.[39] The nonsense variant in *PCDHGA11* (MIM 606298) was validated by Sanger sequencing. The mutation was found to segregate precisely with the *SHANK1* deletion. Because *SHANK1* and *PCDHGA11* reside on different autosomes, we tested for translocation or transposition and ruled out such linkage (Figure S6). It is possible that the Tyr313* mutation in *PCDHGA11* works in concert with the *SHANK1* deletion to modify (positively or negatively) the extent of the phenotype or that they are just randomly cosegregating (a 1/262,144 chance). We have not detected CNVs or sequence-level mutations in *PCDHGA11* in family 2 or in any other ASD subject examined. We have also tentatively ruled out the role of the X chromosome in family 1 given that different X chromosomes were observed in ASD males (by comparing SNP genotypes), and no pathogenic CNV, mutation, or genetic linkage was observed at the X chromosome (Figure S7). The *PCDHGA11* and *SHANK1* loci at 5q31.3 and 19q13.33, respectively, represented two of five chromosomal regions in which maximal linkage was detected (Figure S7). No other obvious damaging sequence mutations or CNVs were found in genes in the other three putative linkage regions.

We provide here a description of hemizygous deletions of *SHANK1* in ASD. Such findings have, perhaps, been anticipated given that mutations leading to haploinsufficiency of *SHANK2* and *SHANK3* have been previously described in ASD. The striking segregation of ASD in only male *SHANK1*-deletion carriers in family 1 represents, to our knowledge, the first example of autosomal sex-limited expression in ASD.[40] Our finding of an unrelated ASD-affected male carrying an independent de novo deletion of *SHANK1* supports our interpretation that the *SHANK1* CNV segregating in family 1 is indeed the primary etiologic event leading to ASD. Additional case reports of other multigenerational families will be required for substantiating our findings of gender-influenced autosomal penetrance differences at the *SHANK1* locus in ASD.

Our data indicate that *SHANK1* deletions are associated with ASD with higher functioning in males. Insofar as all affected males have an IQ in the typical range and have good verbal ability (with a lack of clinically significant language delay), they would also qualify for a diagnosis of Asperger disorder. The one other male (individual I-1) had a diagnosis of the BAP (although this diagnosis was based only on retrospective reports, which preclude the potential for an ASD diagnosis). The female carriers do not show evidence of ASD or the BAP but have suffered from anxiety, which is considered to be an unrelated phenotype to ASD (Table 1). It is surprising that, compared to other studies of inherited CNVs,[3] this pedigree shows little variability in clinical features among affected individuals.

We note that the neuronal genes *PCDHGA11* and *SYT3* could also contribute to aspects of the ASD phenotype in families 1 and 2, respectively. Moreover, subtler mutations,[41,42] or combinations of mutations,[17] might also be involved. The latter possibility might be particularly relevant given early findings that suggest potential multigenic effects including *SHANK2* in ASD risk.[17] Paternal silencing or imprinting of *SHANK1*, as well as potential transmission distortion with respect to cosegregation of *SHANK1* and *PCDHGA11*, also needs to be considered, but at least for the former, there is insufficient evidence for this.[43]

Consistent with our findings in humans, *Shank1*-null murines have deficits in several elements of social communication and developmental milestones.[44] They also exhibit increased anxiety-related behavior and impaired memory.[45] Shank2 and Shank3 offer immature excitatory synapses with unique properties that facilitate synaptogenesis. Shank1 has the important, but perhaps less critical, structural role of "consolidating" novel synaptic contacts via capping Shank2 and Shank3.[22,46] Mutations in *SHANK1* might be predicted generally to have less of an impact on synapse morphology and stability and, therefore, a less severe phenotype than do similar alterations in *SHANK2* or *SHANK3*. Given that *SHANK2* and *SHANK3* mutations are observed in ID[16,38] and schizophrenia (MIM 181500),[47] it is possible that *SHANK1* mutations will also be similarly found in other brain disorders.

There are currently three prevailing theories explaining the gender bias in autism. The "extreme brain" theory proposes that autistic behaviors are an exaggeration of typical male personality traits.[48] A second theory suggests that given the established genetic basis of autism, it is possible that the sex bias lies in the sex chromosomes. Rare mutations in some X-linked genes such as *NLGN3* and *NLGN4*[13] and the *PTCHD1* locus (MIM 300828)[49] have been shown to confer risk to idiopathic ASD in male carriers. A third theory[50] posits a sex-specific multiple-threshold model in which females need to carry more genetic liability than males in order to develop ASD. Some protective factor must exist to account for that higher threshold. Our discovery of rare autosomal *SHANK1* deletions associated with ASD or the BAP that is limited to males provides further evidence of the existence of this protective factor and has significant implications

for elucidating the molecular basis of sex bias in ASD. These results will also have immediate relevance for clinical genetic testing in ASD.

Supplemental Data

Supplemental Data include seven figures and six tables and can be found with this article online at http://www.cell.com/AJHG.

Acknowledgments

This work was supported by grants from the University of Toronto McLaughlin Centre, NeuroDevNet, Genome Canada and the Ontario Genomics Institute, the Canadian Institutes for Health Research (CIHR), the Canadian Institute for Advanced Research, the Canada Foundation for Innovation, the government of Ontario, Autism Speaks, and The Hospital for Sick Children Foundation. S.W.S. holds the GlaxoSmithKline-CIHR Chair in Genome Sciences at the University of Toronto and The Hospital for Sick Children. G.A.R. was supported by Deutsche Forschungsgemeinschaft and the BMBF/NGFNplus (German Mental-Retardation Network). T.B. was supported by Agence Nationale de la Recherche (ANR-08-MNPS-037-01—SynGen), Neuron-ERANET (EUHF-AUTISM), and Fondation Orange. We are indebted to the individuals and their families for participating in this study and to The Centre for Applied Genomics at The Hospital for Sick Children. We thank A. Fiebig, A. Franke, and S. Schreiber at POPGEN (University of Kiel, Kiel, Germany), A. Stewart, R. McPherson, and R. Roberts of the University of Ottawa Heart Institute (University of Ottawa, Ottawa, Canada), the Wellcome Trust Case Control Consortium, and the Study of Addiction: Genetics and Environment consortium for providing control data. A provisional patent has been filed for The Hospital for Sick Children in S.W. Scherer's name.

Received: February 17, 2012
Revised: March 9, 2012
Accepted: March 19, 2012
Published online: April 12, 2012

Web Resources

The URLs for data presented herein are as follows:

Database of Genomic Variants, http://projects.tcag.ca/variation/
dbSNP, http://www.ncbi.nlm.nih.gov/SNP/
Online Mendelian Inheritance in Man (OMIM), http://www.omim.org/

References

1. Losh, M., Childress, D., Lam, K., and Piven, J. (2008). Defining key features of the broad autism phenotype: A comparison across parents of multiple- and single-incidence autism families. Am. J. Med. Genet. B. Neuropsychiatr. Genet. *147B*, 424–433.
2. Gillberg, C., Cederlund, M., Lamberg, K., and Zeijlon, L. (2006). Brief report: "The autism epidemic". The registered prevalence of autism in a Swedish urban area. J. Autism Dev. Disord. *36*, 429–435.
3. Scherer, S.W., and Dawson, G. (2011). Risk factors for autism: Translating genomic discoveries into diagnostics. Hum. Genet. *130*, 123–148.
4. Devlin, B., and Scherer, S.W. (2012). Genetic Architecture in Autism Spectrum Disorder. Curr. Opin. Genet. Dev., in press. 10.1016/j.gde.2012.03.002.
5. Cook, E.H., Jr., and Scherer, S.W. (2008). Copy-number variations associated with neuropsychiatric conditions. Nature *455*, 919–923.
6. Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-Martin, C., Walsh, T., Yamrom, B., Yoon, S., Krasnitz, A., Kendall, J., et al. (2007). Strong association of de novo copy number mutations with autism. Science *316*, 445–449.
7. Levy, D., Ronemus, M., Yamrom, B., Lee, Y.H., Leotta, A., Kendall, J., Marks, S., Lakshmi, B., Pai, D., Ye, K., et al. (2011). Rare de novo and transmitted copy-number variation in autistic spectrum disorders. Neuron *70*, 886–897.
8. Marshall, C.R., Noor, A., Vincent, J.B., Lionel, A.C., Feuk, L., Skaug, J., Shago, M., Moessner, R., Pinto, D., Ren, Y., et al. (2008). Structural variation of chromosomes in autism spectrum disorder. Am. J. Hum. Genet. *82*, 477–488.
9. Pinto, D., Pagnamenta, A.T., Klei, L., Anney, R., Merico, D., Regan, R., Conroy, J., Magalhaes, T.R., Correia, C., Abrahams, B.S., et al. (2010). Functional impact of global rare copy number variation in autism spectrum disorders. Nature *466*, 368–372.
10. Sanders, S.J., Ercan-Sencicek, A.G., Hus, V., Luo, R., Murtha, M.T., Moreno-De-Luca, D., Chu, S.H., Moreau, M.P., Gupta, A.R., Thomson, S.A., et al. (2011). Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. Neuron *70*, 863–885.
11. Toro, R., Konyukh, M., Delorme, R., Leblond, C., Chaste, P., Fauchereau, F., Coleman, M., Leboyer, M., Gillberg, C., and Bourgeron, T. (2010). Key role for gene dosage and synaptic homeostasis in autism spectrum disorders. Trends Genet. *26*, 363–372.
12. Bourgeron, T. (2009). A synaptic trek to autism. Curr. Opin. Neurobiol. *19*, 231–234.
13. Jamain, S., Quach, H., Betancur, C., Råstam, M., Colineaux, C., Gillberg, I.C., Soderstrom, H., Giros, B., Leboyer, M., Gillberg, C., and Bourgeron, T.; Paris Autism Research International Sibpair Study. (2003). Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. Nat. Genet. *34*, 27–29.
14. Szatmari, P., Paterson, A.D., Zwaigenbaum, L., Roberts, W., Brian, J., Liu, X.Q., Vincent, J.B., Skaug, J.L., Thompson, A.P., Senman, L., et al; Autism Genome Project Consortium. (2007). Mapping autism risk loci using genetic linkage and chromosomal rearrangements. Nat. Genet. *39*, 319–328.
15. Vaags, A.K., Lionel, A.C., Sato, D., Goodenberger, M., Stein, Q.P., Curran, S., Ogilvie, C., Ahn, J.W., Drmic, I., Senman, L., et al. (2012). Rare deletions at the neurexin 3 locus in autism spectrum disorder. Am. J. Hum. Genet. *90*, 133–141.
16. Berkel, S., Marshall, C.R., Weiss, B., Howe, J., Roeth, R., Moog, U., Endris, V., Roberts, W., Szatmari, P., Pinto, D., et al. (2010). Mutations in the SHANK2 synaptic scaffolding gene in autism spectrum disorder and mental retardation. Nat. Genet. *42*, 489–491.
17. Leblond, C.S., Heinrich, J., Delorme, R., Proepper, C., Betancur, C., Huguet, G., Konyukh, M., Chaste, P., Ey, E., Rastam, M., et al. (2012). Genetic and Functional Analyses of *SHANK2* Mutations Suggest a Multiple Hit Model of Autism Spectrum Disorders. PLoS Genet. *8*, e1002521.

18. Durand, C.M., Betancur, C., Boeckers, T.M., Bockmann, J., Chaste, P., Fauchereau, F., Nygren, G., Rastam, M., Gillberg, I.C., Anckarsäter, H., et al. (2007). Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. Nat. Genet. *39*, 25–27.
19. Moessner, R., Marshall, C.R., Sutcliffe, J.S., Skaug, J., Pinto, D., Vincent, J., Zwaigenbaum, L., Fernandez, B., Roberts, W., Szatmari, P., and Scherer, S.W. (2007). Contribution of SHANK3 mutations to autism spectrum disorder. Am. J. Hum. Genet. *81*, 1289–1297.
20. Gauthier, J., Spiegelman, D., Piton, A., Lafrenière, R.G., Laurent, S., St-Onge, J., Lapointe, L., Hamdan, F.F., Cossette, P., Mottron, L., et al. (2009). Novel de novo SHANK3 mutation in autistic patients. Am. J. Med. Genet. B. Neuropsychiatr. Genet. *150B*, 421–424.
21. Sheng, M., and Kim, E. (2000). The Shank family of scaffold proteins. J. Cell Sci. *113*, 1851–1856.
22. Grabrucker, A.M., Schmeisser, M.J., Schoen, M., and Boeckers, T.M. (2011). Postsynaptic ProSAP/Shank scaffolds in the cross-hair of synaptopathies. Trends Cell Biol. *21*, 594–603.
23. Risi, S., Lord, C., Gotham, K., Corsello, C., Chrysler, C., Szatmari, P., Cook, E.H., Jr., Leventhal, B.L., and Pickles, A. (2006). Combining information from multiple sources in the diagnosis of autism spectrum disorders. J. Am. Acad. Child Adolesc. Psychiatry *45*, 1094–1103.
24. Hamdan, F.F., Daoud, H., Piton, A., Gauthier, J., Dobrzeniecka, S., Krebs, M.O., Joober, R., Lacaille, J.C., Nadeau, A., Milunsky, J.M., et al. (2011). De novo SYNGAP1 mutations in nonsyndromic intellectual disability and autism. Biol. Psychiatry *69*, 898–901.
25. Lionel, A.C., Crosbie, J., Barbosa, N., Goodale, T., Thiruvahindrapuram, B., Rickaby, J., Gazzellone, M., Carson, A.R., Howe, J.L., Wang, Z., et al. (2011). Rare copy number variation discovery and cross-disorder comparisons identify risk genes for ADHD. Sci Transl Med. *3*, 95ra75.
26. Pinto, D., Darvishi, K., Shi, X., Rajan, D., Rigler, D., Fitzgerald, T., Lionel, A.C., Thiruvahindrapuram, B., Macdonald, J.R., Mills, R., et al. (2011). Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants. Nat. Biotechnol. *29*, 512–520.
27. Shaikh, T.H., Gai, X., Perin, J.C., Glessner, J.T., Xie, H., Murphy, K., O'Hara, R., Casalunovo, T., Conlin, L.K., D'Arcy, M., et al. (2009). High-resolution mapping and analysis of copy number variations in the human genome: A data resource for clinical and research applications. Genome Res. *19*, 1682–1690.
28. Itsara, A., Cooper, G.M., Baker, C., Girirajan, S., Li, J., Absher, D., Krauss, R.M., Myers, R.M., Ridker, P.M., Chasman, D.I., et al. (2009). Population analysis of large copy number variants and hotspots of human genetic disease. Am. J. Hum. Genet. *84*, 148–161.
29. Krawczak, M., Nikolaus, S., von Eberstein, H., Croucher, P.J., El Mokhtari, N.E., and Schreiber, S. (2006). PopGen: Population-based recruitment of patients and controls for the analysis of complex genotype-phenotype relationships. Community Genet. *9*, 55–61.
30. Stewart, A.F., Dandona, S., Chen, L., Assogba, O., Belanger, M., Ewart, G., LaRose, R., Doelle, H., Williams, K., Wells, G.A., et al. (2009). Kinesin family member 6 variant Trp719Arg does not associate with angiographically defined coronary artery disease in the Ottawa Heart Genomics Study. J. Am. Coll. Cardiol. *53*, 1471–1472.
31. Zogopoulos, G., Ha, K.C., Naqib, F., Moore, S., Kim, H., Montpetit, A., Robidoux, F., Laflamme, P., Cotterchio, M., Greenwood, C., et al. (2007). Germ-line DNA copy number variation frequencies in a large North American population. Hum. Genet. *122*, 345–353.
32. Altshuler, D.M., Gibbs, R.A., Peltonen, L., Altshuler, D.M., Gibbs, R.A., Peltonen, L., Dermitzakis, E., Schaffner, S.F., Yu, F., Peltonen, L., et al; International HapMap 3 Consortium. (2010). Integrating common and rare genetic variation in diverse human populations. Nature *467*, 52–58.
33. Craddock, N., Hurles, M.E., Cardin, N., Pearson, R.D., Plagnol, V., Robson, S., Vukcevic, D., Barnes, C., Conrad, D.F., Giannoulatou, E., et al; Wellcome Trust Case Control Consortium. (2010). Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. Nature *464*, 713–720.
34. Bierut, L.J., Agrawal, A., Bucholz, K.K., Doheny, K.F., Laurie, C., Pugh, E., Fisher, S., Fox, L., Howells, W., Bertelsen, S., et al; Gene, Environment Association Studies Consortium. (2010). A genome-wide association study of alcohol dependence. Proc. Natl. Acad. Sci. USA *107*, 5082–5087.
35. Iafrate, A.J., Feuk, L., Rivera, M.N., Listewnik, M.L., Donahoe, P.K., Qi, Y., Scherer, S.W., and Lee, C. (2004). Detection of large-scale variation in the human genome. Nat. Genet. *36*, 949–951.
36. Zhang, J., Feuk, L., Duggan, G.E., Khaja, R., and Scherer, S.W. (2006). Development of bioinformatics resources for display and analysis of copy number and other structural variants in the human genome. Cytogenet. Genome Res. *115*, 205–214.
37. Park, H., Kim, J.I., Ju, Y.S., Gokcumen, O., Mills, R.E., Kim, S., Lee, S., Suh, D., Hong, D., Kang, H.P., et al. (2010). Discovery of common Asian copy number variants using integrated high-resolution array CGH and massively parallel DNA sequencing. Nat. Genet. *42*, 400–405.
38. Hamdan, F.F., Gauthier, J., Araki, Y., Lin, D.T., Yoshizawa, Y., Higashi, K., Park, A.R., Spiegelman, D., Dobrzeniecka, S., Piton, A., et al; S2D Group. (2011). Excess of de novo deleterious mutations in genes associated with glutamatergic systems in nonsyndromic intellectual disability. Am. J. Hum. Genet. *88*, 306–316.
39. Phillips, G.R., Tanaka, H., Frank, M., Elste, A., Fidler, L., Benson, D.L., and Colman, D.R. (2003). Gamma-protocadherins are targeted to subsets of synapses and intracellular organelles in neurons. J. Neurosci. *23*, 5096–5104.
40. Zhao, X., Leotta, A., Kustanovich, V., Lajonchere, C., Geschwind, D.H., Law, K., Law, P., Qiu, S., Lord, C., Sebat, J., et al. (2007). A unified genetic theory for sporadic and inherited autism. Proc. Natl. Acad. Sci. USA *104*, 12831–12836.
41. Durand, C.M., Perroy, J., Loll, F., Perrais, D., Fagni, L., Bourgeron, T., Montcouquiol, M., and Sans, N. (2012). SHANK3 mutations identified in autism lead to modification of dendritic spine morphology via an actin-dependent mechanism. Mol. Psychiatry *17*, 71–84.
42. Berkel, S., Tang, W., Treviño, M., Vogt, M., Obenhaus, H.A., Gass, P., Scherer, S.W., Sprengel, R., Schratt, G., and Rappold, G.A. (2012). Inherited and de novo SHANK2 variants associated with autism spectrum disorder impair neuronal morphogenesis and physiology. Hum. Mol. Genet. *21*, 344–357.
43. Gregg, C., Zhang, J., Weissbourd, B., Luo, S., Schroth, G.P., Haig, D., and Dulac, C. (2010). High-resolution analysis of parent-of-origin allelic expression in the mouse brain. Science *329*, 643–648.

44. Wöhr, M., Roullet, F.I., Hung, A.Y., Sheng, M., and Crawley, J.N. (2011). Communication impairments in mice lacking Shank1: Reduced levels of ultrasonic vocalizations and scent marking behavior. PLoS ONE *6*, e20631.
45. Hung, A.Y., Futai, K., Sala, C., Valtschanoff, J.G., Ryu, J., Woodworth, M.A., Kidd, F.L., Sung, C.C., Miyakawa, T., Bear, M.F., et al. (2008). Smaller dendritic spines, weaker synaptic transmission, but enhanced spatial learning in mice lacking Shank1. J. Neurosci. *28*, 1697–1708.
46. Grabrucker, A.M., Knight, M.J., Proepper, C., Bockmann, J., Joubert, M., Rowan, M., Nienhaus, G.U., Garner, C.C., Bowie, J.U., Kreutz, M.R., et al. (2011). Concerted action of zinc and ProSAP/Shank in synaptogenesis and synapse maturation. EMBO J. *30*, 569–581.
47. Gauthier, J., Champagne, N., Lafrenière, R.G., Xiong, L., Spiegelman, D., Brustein, E., Lapointe, M., Peng, H., Côté, M., Noreau, A., et al; S2D Team. (2010). De novo mutations in the gene encoding the synaptic scaffolding protein SHANK3 in patients ascertained for schizophrenia. Proc. Natl. Acad. Sci. USA *107*, 7863–7868.
48. Baron-Cohen, S., Lombardo, M.V., Auyeung, B., Ashwin, E., Chakrabarti, B., and Knickmeyer, R. (2011). Why are autism spectrum conditions more prevalent in males? PLoS Biol. *9*, e1001081.
49. Noor, A., Whibley, A., Marshall, C.R., Gianakopoulos, P.J., Piton, A., Carson, A.R., Orlic-Milacic, M., Lionel, A.C., Sato, D., Pinto, D., et al. (2010).). Disruption at the PTCHD1 Locus on Xp22.11 in Autism spectrum disorder and intellectual disability. Sci. Transl. Med. *2*, 49ra68.
50. Szatmari, P., Liu, X.Q., Goldberg, J., Zwaigenbaum, L., Paterson, A.D., Woodbury-Smith, M., Georgiades, S., Duku, E., and Thompson, A. (2012). Sex differences in repetitive stereotyped behaviors in autism: Implications for genetic liability. Am. J. Med. Genet. B. Neuropsychiatr. Genet. *159B*, 5–12.

OPEN ACCESS Freely available online PLoS GENETICS

Genetic and Functional Analyses of *SHANK2* Mutations Suggest a Multiple Hit Model of Autism Spectrum Disorders

Claire S. Leblond[1,2,3], Jutta Heinrich[4], Richard Delorme[1,2,5], Christian Proepper[4], Catalina Betancur[6,7,8], Guillaume Huguet[1,2,3], Marina Konyukh[1,2,3], Pauline Chaste[1,2,3], Elodie Ey[1,2,3], Maria Rastam[9], Henrik Anckarsäter[10], Gudrun Nygren[11], I. Carina Gillberg[11], Jonas Melke[12], Roberto Toro[1,2,3], Beatrice Regnault[13], Fabien Fauchereau[1,2,3], Oriane Mercati[1,2,3], Nathalie Lemière[1,2,3], David Skuse[14], Martin Poot[15], Richard Holt[16], Anthony P. Monaco[16], Irma Järvelä[17], Katri Kantojärvi[17], Raija Vanhala[17], Sarah Curran[18], David A. Collier[19], Patrick Bolton[18,19], Andreas Chiocchetti[20], Sabine M. Klauck[20], Fritz Poustka[21], Christine M. Freitag[21], Regina Waltes[21], Marnie Kopp[21], Eftichia Duketis[21], Elena Bacchelli[22], Fiorella Minopoli[22], Liliana Ruta[23], Agatino Battaglia[24], Luigi Mazzone[25], Elena Maestrini[22], Ana F. Sequeira[26,27,28], Barbara Oliveira[26,27,28], Astrid Vicente[26,27,28], Guiomar Oliveira[29], Dalila Pinto[30], Stephen W. Scherer[30], Diana Zelenika[31], Marc Delepine[31], Mark Lathrop[31], Dominique Bonneau[32,33], Vincent Guinchat[34], Françoise Devillard[35], Brigitte Assouline[34], Marie-Christine Mouren[5], Marion Leboyer[36,37,38], Christopher Gillberg[11,39], Tobias M. Boeckers[4], Thomas Bourgeron[1,2,3]*

1 Human Genetics and Cognitive Functions, Institut Pasteur, Paris, France, **2** CNRS URA 2182 "Genes, synapses and cognition," Institut Pasteur, Paris, France, **3** University Denis Diderot Paris 7, Paris, France, **4** Institute of Anatomy and Cell Biology, Ulm University, Ulm, Germany, **5** Assistance Publique-Hôpitaux de Paris, Robert Debré Hospital, Department of Child and Adolescent Psychiatry, Paris, France, **6** INSERM, U952, Paris, France, **7** CNRS, UMR 7224, Paris, France, **8** UPMC Univ Paris 06, Paris, France, **9** Department of Clinical Sciences in Lund, Lund University, Lund, Sweden, **10** Institute of Clinical Sciences, Lund University, Malmö, Sweden, **11** Gillberg Neuropsychiatry Centre, University of Gothenburg, Göteborg, Sweden, **12** Institute of Neuroscience and Physiology, Department of Pharmacology, Gothenburg University, Göteborg, Sweden, **13** Eukaryote Genotyping Platform, Genopole, Institut Pasteur, Paris, France, **14** Behavioural and Brain Sciences Unit, Institute of Child Health, University College London, London, United Kingdom, **15** Department of Medical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands, **16** Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom, **17** Department of Medical Genetics, University of Helsinki, Helsinki, Finland, **18** Academic Department of Child and Adolescent Psychiatry, Institute of Psychiatry, King's College London, London, United Kingdom, **19** Social Genetic Developmental Psychiatry Centre, Institute of Psychiatry, King's College London, London, United Kingdom, **20** Division of Molecular Genome Analysis, German Cancer Research Center (DKFZ), Heidelberg, Germany, **21** Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, Goethe University, Frankfurt am Main, Germany, **22** Department of Biology, University of Bologna, Bologna, Italy, **23** Division of Child Neurology and Psychiatry, Department of Paediatrics, University of Catania, Catania, Italy, **24** Stella Maris Clinical Research Institute for Child and Adolescent Neuropsychiatry, Pisa, Italy, **25** Division of Child Neurology and Psychiatry, Department of Pediatrics, University of Catania, Catania, Italy, **26** Instituto Nacional de Saude Dr Ricardo Jorge, Lisbon, Portugal, **27** Instituto Gulbenkian de Ciencia, Oeiras, Portugal, **28** Center for Biodiversity, Functional and Integrative Genomics, Faculdade de Ciências da Universidade de Lisboa, Lisboa, Portugal, **29** Unidade Neurodesenvolvimento e Autismo, Centro Investigação e Formação Clinica, Hospital Pediátrico Coimbra e Faculdade Medicina, Universidade Coimbra, Coimbra, Portugal, **30** The Centre for Applied Genomics and Program in Genetics and Genomic Biology, The Hospital for Sick Children, Toronto, Canada, **31** Centre National de Génotypage, Evry, France, **32** INSERM U771 and CNRS 6214, Angers, France, **33** Département de Biochimie et Génétique, Centre Hospitalier Universitaire, Angers, France, **34** CADIPA–Centre de Ressources Autisme Rhône-Alpes, Saint Egrève, France, **35** Genetics Department, Hôpital Couple-Enfant, Grenoble, France, **36** INSERM, U955, Psychiatrie Génétique, Créteil, France, **37** Université Paris Est, Faculté de Médecine, Créteil, France, **38** AP-HP, Hôpital H. Mondor–A. Chenevier, Département de Psychiatrie, Créteil, France, **39** Institute of Child Health, University College London, London, United Kingdom

Abstract

Autism spectrum disorders (ASD) are a heterogeneous group of neurodevelopmental disorders with a complex inheritance pattern. While many rare variants in synaptic proteins have been identified in patients with ASD, little is known about their effects at the synapse and their interactions with other genetic variations. Here, following the discovery of two *de novo SHANK2* deletions by the Autism Genome Project, we identified a novel 421 kb *de novo SHANK2* deletion in a patient with autism. We then sequenced *SHANK2* in 455 patients with ASD and 431 controls and integrated these results with those reported by Berkel *et al.* 2010 (n = 396 patients and n = 659 controls). We observed a significant enrichment of variants affecting conserved amino acids in 29 of 851 (3.4%) patients and in 16 of 1,090 (1.5%) controls (P = 0.004, OR = 2.37, 95% CI = 1.23–4.70). In neuronal cell cultures, the variants identified in patients were associated with a reduced synaptic density at dendrites compared to the variants only detected in controls (P = 0.0013). Interestingly, the three patients with *de novo SHANK2* deletions also carried inherited CNVs at 15q11–q13 previously associated with neuropsychiatric disorders. In two cases, the nicotinic receptor *CHRNA7* was duplicated and in one case the synaptic translation repressor *CYFIP1* was deleted. These results strengthen the role of synaptic gene dysfunction in ASD but also highlight the presence of putative modifier genes, which is in keeping with the "multiple hit model" for ASD. A better knowledge of these genetic interactions will be necessary to understand the complex inheritance pattern of ASD.

Citation: Leblond CS, Heinrich J, Delorme R, Proepper C, Betancur C, et al. (2012) Genetic and Functional Analyses of *SHANK2* Mutations Suggest a Multiple Hit Model of Autism Spectrum Disorders. PLoS Genet 8(2): e1002521. doi:10.1371/journal.pgen.1002521

Editor: Matthew State, Yale University School of Medicine, United States of America

Received March 23, 2011; **Accepted** December 11, 2011; **Published** February 9, 2012

Funding: This work was supported by the Institut Pasteur, INSERM, APHP, ANR (ANR-08-MNPS-037-01—SynGen), Neuron-ERANET (EUHF-AUTISM), Fondation Orange, Fondation pour la Recherche Médicale, RTRS Santé Mentale (Foundation FondaMental), the Deutsche Forschungsgemeinschaft DFG (BO1718:3-1 and SFB497/B8), the Dutch Foundation for Brain Research (Hersenstichting, grant # 2008(1).34), and Wellcome Trust core grant (075491/Z/04). D Pinto is supported by a postdoctoral fellowship from the Canadian Institutes of Health Research (#213997). SW Scherer holds the GlaxoSmithKline-CIHR Pathfinder Chair in Genetics and Genomics at the University of Toronto and the Hospital for Sick Children. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: thomasb@pasteur.fr

Introduction

Autism spectrum disorders (ASD) are characterized by impairments in reciprocal social communication and stereotyped behaviors [1]. The prevalence of ASD is about 1/100, but closer to 1/300 for typical autism [2]. ASD are more common in males than females, with a 4:1 ratio. Previously, twin and family studies have conclusively described ASD as the most ''genetic'' of neuropsychiatric disorders, with concordance rates of 82–92% in monozygotic twins versus 1–10% in dizygotic twins [3], but a recent study finds evidence for a more substantial environmental component [4]. In the absence of Mendelian inheritance patterns, ASD were first considered to be polygenic, i.e., a disorder caused by multiple genetic risk factors, each of weak effect. More recently, an alternative model was proposed that considered ASD as a group of disorders caused by heterogeneous genetic risk factors influencing common neuronal pathways [5,6]. It was supported by the identification of apparently monogenic forms of ASD, each affecting a limited number of patients (1–2% for the most replicated genes) [7–14]. In this model, eventually a single highly penetrant mutation would be sufficient to produce ASD. However, the occurrence of two or more deleterious copy number variants (CNV) or mutations in a subset of patients also suggested that independent loci could act in concert to induce the development of ASD [9,13–16]. In line with these findings, the recent observation that patients with a deletion at 16p12.1 were more likely to carry an additional large CNV agrees with a ''two-hit model'' for developmental disorders [17].

The genetic causes of ASD are diverse [18], but the main category of genes associated with the disorder is related to the development and function of neuronal circuits [6,19]. Mutations of genes coding for synaptic cell adhesion molecules and scaffolding proteins, such as neuroligins (NLGN), neurexins (NRXN) and SHANK, have been recurrently reported in patients with ASD [7–10,13,14,20]. These proteins play a crucial role in the formation and stabilization of synapses [21], as well as in synaptic homeostasis [22]. *SHANK2* and *SHANK3* code for scaffolding proteins located in the postsynaptic density (PSD) of glutamatergic synapses. Deletions of *ProSAP2/SHANK3* at chromosome 22q13 are one of the major genetic abnormalities in neurodevelopmental disorders [20], and mutations of *ProSAP2/SHANK3* have been identified in patients with ASD, intellectual disability (ID) and schizophrenia [7,23–25]. Mutations of *ProSAP1/SHANK2* have also recently been reported in both, ASD and ID [9,26]. The difference in clinical outcome of mutation carriers has been attributed to the presence of still uncharacterized additional genetic, epigenetic and/or environmental factors [27].

In order to better understand the role of the NRXN-NLGN-SHANK pathway in ASD, we first aimed to describe *SHANK2* isoform expression in different tissues of healthy individuals. To investigate the role of this pathway in ASD, we screened for *SHANK2* CNVs and coding mutations in a large sample of patients with ASD and controls. We provide genetic and functional evidence that *SHANK2* is associated with ASD, and that its mutations affect the number of synapses. Additionally, we report the co-occurrence of *SHANK2 de novo* deletions and inherited CNVs altering neuronal genes, suggesting that epistasis between specific loci in the genome could modulate the risk for ASD.

Results

SHANK2 isoforms are differentially expressed in human tissues

In order to characterize all isoforms of *SHANK2*, we scanned genomic databases for specific Expressed Sequence Tags (ESTs) and spliced isoforms. The human *SHANK2* gene (*NM_012309.3*) spans 621.8 kb and contains 25 exons (Figure 1). The longest *SHANK2* isoform (*SHANK2E*, *AB208025*) contains ankyrin (ANK) repeats at the N-terminus, followed by a Src homology 3 (SH3) domain, a PSD95/DLG/ZO1 (PDZ) domain, a proline-rich region and a sterile alpha motif (SAM) domain at its C-terminus region. All these domains are involved in protein-protein interactions that bridge glutamate receptors, scaffolding proteins and intracellular effectors to the actin cytoskeleton [28,29]. Two additional isoforms, *ProSAP1A* (*AB208026*) and *ProSAP1* (*AB208027*), originating from distinct promoters, were previously detected in the rat [30,31]. Finally, the shortest isoform (*AF141901*), also originally described in the rat, results in premature termination of the transcription before the SAM domain due to an alternative 3′ end in exon 22 [32] (Figure 1A). To validate these *SHANK2* isoforms in humans, we used specific RT-PCRs and sequencing (Figure 1B). Almost all tissues tested (brain, liver, placenta, kidney, lung, pancreas and lymphoblastoid cell lines) expressed *SHANK2* mRNA, except heart and skeletal muscle, for which no expression was detected. We observed inter-individual differences in the relative amount of *SHANK2* mRNA that were confirmed by using independent RT-PCRs and primers (not shown). Such differences have been previously reported for other synaptic genes such as *NLGN1-4Y*, *PCHD11X/Y*, and *SHANK3* [7,8,33] and might be the consequence of polymorphisms located in specific regulatory sequences and/or activity dependent expression of this family of post-synaptic proteins [34]. Notably, exons 19, 20 and 23 were found to be expressed only in brain in all individuals tested (Figure 1C). Such brain specific splicing has been already observed for exon 18 in *SHANK3* [7], which is similar to exon 19 and 20 in *SHANK2*. These 'brain-

Author Summary

Autism spectrum disorders (ASD) are a heterogeneous group of neurodevelopmental disorders with a complex inheritance pattern. While mutations in several genes have been identified in patients with ASD, little is known about their effects on neuronal function and their interaction with other genetic variations. Using a combination of genetic and functional approaches, we identified novel *SHANK2* mutations including a *de novo* loss of one copy of the *SHANK2* gene in a patient with autism and several mutations observed in patients that reduced neuronal cell contacts *in vitro*. Further genomic analysis of three patients carrying *de novo SHANK2* deletions identified additional rare genomic imbalances previously associated with neuropsychiatric disorders. Taken together, these results strengthen the role of synaptic gene dysfunction in ASD but also highlight the presence of putative modifier genes, which is in keeping with the "multiple hit model" for ASD. A better knowledge of these genetic interactions will be necessary to understand the complex inheritance pattern of ASD.

specific exons' code for a region in SHANK2/3 located between the PDZ and the proline rich domains. Finally, in contrast to previous results [26], we detected the longest *SHANK2E* isoform in all independent samples of human brain, with high expression in the cerebellum (Figure 1, Figure S1). This *Shank2E* isoform was also expressed in the cerebellum and in the liver of rat embryo at E19 (Figure S1).

A *de novo* deletion of *SHANK2* in a patient with ASD

Berkel *et al.* 2010 recently identified two independent *de novo SHANK2* deletions in two patients, one with ID and another one with ASD [26]. In addition, whole genome analysis performed by the Autism Genome Project (AGP) using Illumina 1M single nucleotide polymorphism (SNP) arrays detected one additional *de novo SHANK2* deletion in a patient (6319_3) with ASD [9] (the second patient described by the AGP, 5237_3, is patient SK0217-003 reported in Berkel *et al.* 2010 [26]). Recently, a 3.4 Mb de novo deletion including *SHANK2* was observed in a female patient with speech and developmental delay [35]. To follow up on these results, we genotyped an independent sample of 260 patients with ASD using Illumina 1M Duo SNP arrays (Table S1). In this sample, we detected a 421.2 kb deletion within *SHANK2* in patient AU038_3 with autism and moderate ID (see patient section in Materials and Methods, and Table S2). The deletion covered twelve exons (E5–E16) and altered all *SHANK2* isoforms (Figure 2A). No other deleterious variants in the remaining copy of *SHANK2* were detected by sequencing. The parents did not carry the deletion, indicating a *de novo* event. The deletion was validated by quantitative PCR analysis using DNA from an independent blood sample from all members of the family and SNP analysis indicated that the deletion originated on the maternal chromosome (Figure S2). *SHANK2* deletions were absent in more than 5000 controls [9,26] and not listed in the Database of Genomic Variants (DGV; http://projects.tcag.ca/variation/).

SHANK2 coding variants affecting conserved amino acids are enriched in patients with ASD

To probe for additional mutations, we first sequenced all exons of the longest *SHANK2E* isoform in 230 patients with ASD and 230 controls. We then sequenced an additional sample of 225 patients and 201 controls (Table S1) for the *ProSAP1A* isoform that corresponds to the major *SHANK2* isoform in the brain. Since we screened all *SHANK2* isoforms, we used a nomenclature including the *SHANK2E* isoform that differed from Berkel *et al.* 2010 [26]. Within the 9 coding exons specific to *SHANK2E*, we identified R174C (rs7926203) listed in dbSNP in 2 independent patients with ASD and R185Q in one patient with ASD. For this isoform, no variant was identified in the control sample. Within the *ProsSAP1A* isoform, we identified 24 non-synonymous variations. When these results are integrated with those obtained by Berkel *et al.* 2010, a total of 40 variants of *ProsSAP1A* including 3 already reported in dbSNP were identified (Figure 2B, Table 1, Figure S3). Only two variants (Y967C and R569H) with MAF>1% are detected and there is no enrichment of rare variants of *SHANK2* (MAF<1%) in patients with ASD compared with controls. Because variants affecting conserved amino acids in the SHANK proteins are most likely to have a functional effect, we tested whether there was an enrichment of these variants in patients compared to controls. The alignment of the SHANK protein sequences and the conservation of the variants are indicated in the Table S5. In both mutation screening studies, the first performed by Berkel *et al.* 2010 and the second presented here, we observed an enrichment of variants affecting conserved amino acids in patients compared with controls (Figure 2C, Table S5 and Table S7). Overall, 12 of 15 (80%) of the variants identified only in the patient sample affected conserved amino acids compared with only 6 of 17 (35.3%) in controls (Fisher's exact test 1-sided, P = 0.013, OR = 6.83, 95% IC = 1.19–53.40). Because several independent patients carried these variants (Table 1), the enrichment is even more significant when the number of carriers was considered. The variants affecting conserved amino acids were observed in 29 of 851 (3.4%) patients and in 16 of 1090 (1.5%) controls (Fisher's exact test 1-sided, P = 0.004, OR = 2.37, 95% CI = 1.23–4.70). A total of 8 variants were identified in patients and controls. Among these 8 variants, 2 affected conserved amino acids (R818H and S557N). The variant S557N was observed in 9 of 851 (1.06%) independent families with ASD and in 3 of 1090 (0.28%) controls (Fisher's Exact Test one sided, P = 0.029, OR = 3.87; 95% CI = 0.96–22.29). It affects a conserved serine with a high probability of being phosphorylated and located in the SH3 domain of all SHANK proteins. This domain binds to GRIP and b-PIX, two proteins linking SHANK to glutamate AMPA receptors and actin skeleton, respectively [36]. In our initial mutation screen, R818H was observed in 5 of 230 patients with ASD and 0 of 230 controls. In order to determine if R818H was more frequent in the patients with ASD, we screened an additional sample of 3020 individuals with ASD, 1783 controls from European descent, and the Human Genome Diversity Panel (HGDP) control dataset (Table S3 and S4). R818H was virtually absent outside Europe and had the highest allelic frequency (2.37%) in Finland, but overall its frequency was not higher in patients with ASD compared with controls (ASD 32/3250 (1.0%); controls 27/2030 (1.33%); Fisher's exact test 2-sided, P = 0.28) (Table S3).

Finally, and unexpectedly, during this additional mutation screening, we detected a variation (IVS22+1G>T) altering the consensus donor splice site of exon 22 in a Swedish control, SWE_Q56_508 (Figure 3A). This variant was predicted to disrupt all *SHANK2* isoforms by deleting the proline rich and the SAM domain, except for the shortest isoform *AF141901*, where the mutation is located in the open reading frame (ORF) and should lead to a G263V change. This variant was not observed in 1786 patients or 1407 controls, and is not listed in dbSNP. This control female was part of a previous epidemiological study [37] and had been extensively examined for anthropometrics and cardiovascu-

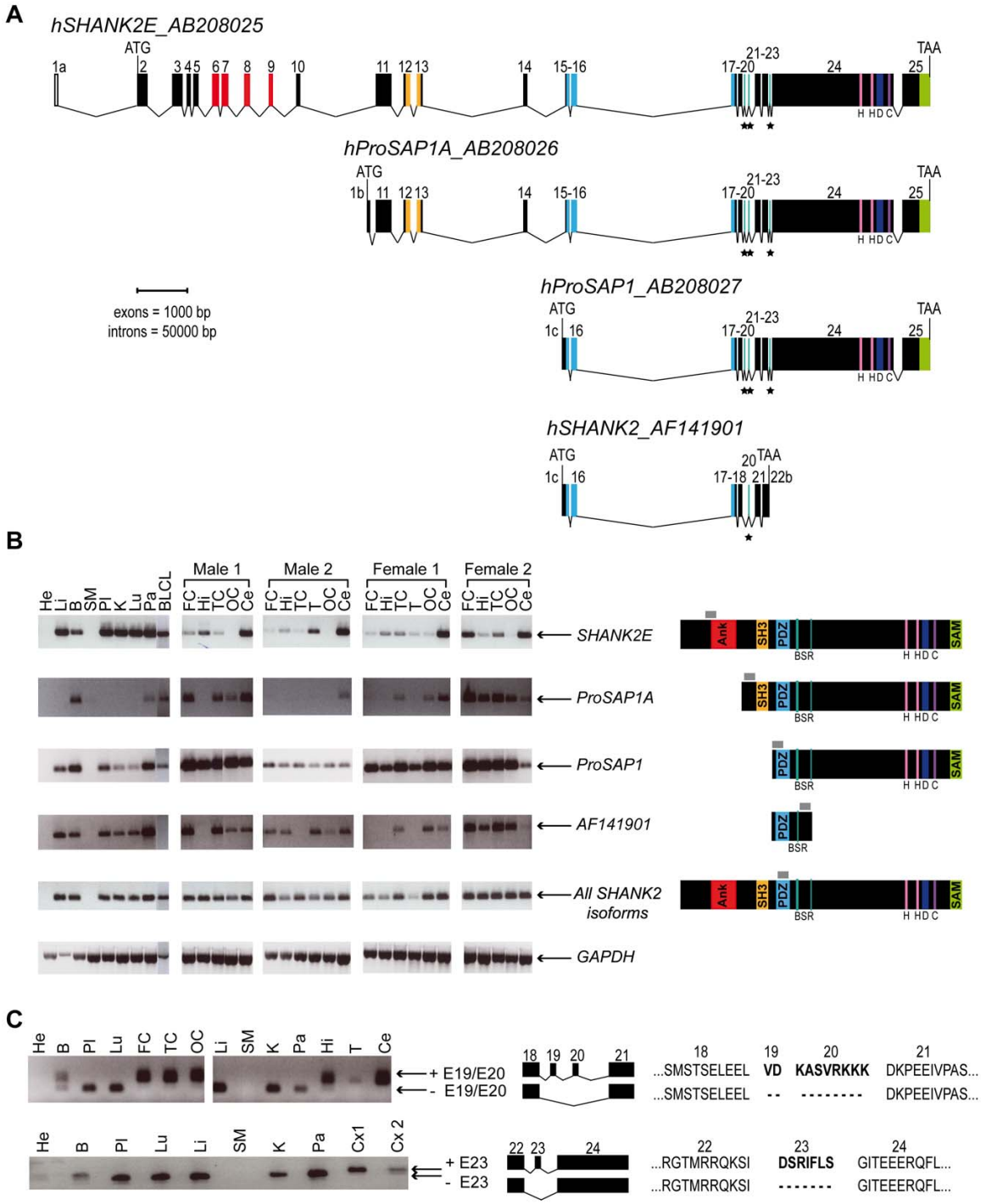


Figure 1. Genomic structure, isoforms, and expression of human *SHANK2*. A. Genomic structure of the human *SHANK2* gene. Transcription of *SHANK2* produces four main mRNA from three distinct promoters: *SHANK2E* (*AB208025*), *ProSAP1A* (*AB208026*), *ProSAP1* (*AB208027*) and *AF141901*. There are three translation starts: in exon 2 for *SHANK2E*, in exon1b for *ProSAP1A*, and in exon1c for *ProSAP1* and *AF141901*; and two independent stop codons: in exon 22b for *AF141901* and in exon 25 for *SHANK2E*, *ProSAP1A* and *ProSAP1*. Conserved domains of protein interaction or protein binding site are represented in color: ANK (red), SH3 (orange), PDZ (blue) and SAM (green), H (pink), D, (dark blue) and C (purple). Black stars identify the alternative spliced exons ('brain-specific exons' in turquoise: 19, 20 and 23). B. RT-PCRs of *SHANK2* isoforms on RNA from different human control tissues (Clontech), and different brain regions of four controls (2 males and 2 females). The amplified regions specific to each isoform of *SHANK2* are

indicated by gray boxes. C. Alternative splicing of human *SHANK2*; exons 19, 20 and 23 are specific to the brain. ANK, ankyrin; SH3, Src homology 3; PDZ, PSD95/DLG/ZO1; SAM, sterile alpha motif; He, heart; Li, liver; B, brain; SM, skeletal muscle; Pl, placenta; K, kidney; Lu, lung; Pa, pancreas; FC, frontal cortex; Hi, hippocampus; TC, temporal cortex; T, thalamus; OC, occipital cortex; Ce, cerebellum; Cx, whole cortex; BLCL, B lymphoblastoid cell lines; GAPDH, glyceraldehyde 3-phosphate dehydrogenase; BSR, brain specific region; H, homer binding site; D, dynamin binding site; C, cortactin binding site. The ages of the two males and the two females studied were 74, 42, 55, and 36 years with a post-mortem interval of 10, 21, 24, and 2 h, respectively.
doi:10.1371/journal.pgen.1002521.g001

lar risk factors such as blood pressure and levels of all major hormones. In addition, she was ascertained for axis I psychiatric disorders and personality traits using the Temperament and Character Inventory (TCI) [38] and the Karolinska Scales of Personality (KSP) [39]. Notably, despite the predicted deleterious effect of the mutation, this subject had no major somatic or psychiatric health problems. Regarding personality traits, none of her scores for TCI items were different from those found in the general population. KSP assessment showed that her scores for neuroticism (51.3), nonconformity/aggressiveness (56.7), and psychoticism (50.5) were not different from the general population (mean $\pm$ SD = 50$\pm$10). However, she displayed a high score (61.4) for the extraversion factor, and for one of its subscales, monotony avoidance [40].

Several *SHANK2* variants identified in patients alter synapse density in cultured neurons

In order to establish the functional impact of *SHANK2* variations, we performed expression studies in primary neuronal cell cultures after over-expression of wild-type *vs.* mutant *ProSAP1A/Shank2A* cDNA (Figure 4). All the variants (n = 16) identified by our first screen of 230 patients and 230 controls were tested: 5 were identified only in patients (V717F, A729T, G1170R, D1535N and L1722P), 6 were detected in patients and controls (S557N, R569H, K780Q, R818H, Y967C and P1586L) and 5 were only found in controls (L629P, A822T, V823M, R1290W and Q1308R). All variants were predicted as damaging by Polyphen2 DIV except Q1308R identified only in controls and predicted as benign [41]. In the patient sample, 5/5 variants affected conserved amino acids in the SHANK proteins compared with only 2/6 in the group of variants identified in patients and controls, and 2/5 in the control group. All mutation sites were introduced into the rat *ProSAP1A* cDNA and confirmed by sequencing. The effect of the *Shank2* variants was further investigated in cultured hippocampal neurons. Upon transfection, Western blot analysis revealed that the different GFP-Shank2 fusion proteins were expressed with the expected size (Figure S4). Results from quantification showed that none of the variants affected the cluster formation of Shank2 protein along the dendrites, the number or the general branching pattern of dendrites (Figure S4). In contrast, 8 variants identified in patients or in patients and control group reduced significantly the density of Shank2 positive synapses per 10 µm dendrite length compared with wild-type GFP-Shank2 (Figure 4). None of the variants identified in controls only were shown to have a significant effect. After Bonferonni correction for the 16 tests, 4 variants significantly affected synapse density. Among these variants, A729T, G1170R and L1722P were identified only in patients and S557N was observed more frequently in patients than in controls. As expected, the majority of the variants leading to reduced synaptic density altered conserved amino acids present in SHANK proteins (7/8), and the majority of variants not changing synaptic density affected amino acids present only in SHANK2 (7/8). The 4 strongly associated (after Bonferroni correction) variants affecting synaptic density modified conserved amino acids in other SHANK proteins. Because the significant threshold of 0.05 is arbitrary, we additionally tested for the quantitative effect of the variant on synaptic density as a continuous trait (Figure S4) and found that variants identified in patients were associated with a significant decrease of synapse density *in vitro* compared with those shared by patients and controls (Student's t test 2-sided P = 0.022) or those only detected in the controls (Student's t test 2-sided P = 0.0013). As expected, variants affecting conserved amino acids were associated with a higher reduction of synapse density *in vitro* (Student's t test 2-sided P = 0.014).

Additional CNVs affect neuronal genes in patients with *de novo SHANK2* deletions and in the control carrying the *SHANK2* splice mutation

To test if additional CNVs may modulate the impact of *SHANK2* mutations in the development of ASD, we analyzed the CNVs of patient AU038_3 and the two patients (5237_3 and 6319_3) carrying *SHANK2 de novo* deletions previously identified by Pinto et al. [9] (Figure 5 and Table S6). In addition to our CNV study group of 260 patients with ASD and 290 controls, we used the CNV dataset from the AGP, which includes 996 patients with ASD and 1287 controls genotyped with the Illumina 1M SNP array [9]. Remarkably, all three patients with *SHANK2 de novo* deletions also carried rare inherited genetic imbalances at chromosome 15q11–q13 (Figure 6), a region associated with Angelman syndrome, Prader-Willi syndrome and other neuropsychiatric disorders, including ASD [42–61]. This region is characterized by recurrent deletions/duplications with breakpoints generally located within five segmental duplications named BP1 to BP5, which act as hotspots of non-allelic homologous recombination. In the BP5 region, patients AU038_3 and 5237_3 carried the same 496 kb duplication of the nicotinic receptor *CHRNA7* gene (29.8–30.3 Mb, hg 18; maternally inherited in patient AU038_3 and paternally inherited in patient 5237_3). This small *CHRNA7* duplication was present in 13 of 1257 patients with ASD (1.03%) compared with 9 of 1577 controls (0.57%) (Fisher's exact test, 2-sided P = 0.19).

These duplications are considered of uncertain clinical significance since they were previously detected at similar frequencies in patients with epilepsy (6 of 647, 0.93%), in controls (19 of 3699, 0.51%) [50], and in subjects referred for chromosomal microarray analysis (55 of 8832, 0.62%) [51]. In contrast, larger 15q13.3 deletions (~1.5 Mb) between BP4 and BP5, encompassing the *CHRNA7* locus have been associated with disorders such as ID, epilepsy, schizophrenia, and ASD [43,46–48,50,52–54,57–59]. In the BP4 region, the same two patients AU038_3 and 5237_3 also carried two independent deletions of the rhoGAP *ARHGAP11B* gene. Loss of *ARHGAP11B* was detected in 8 of 1257 patients with ASD (0.64%) and in 4 of 1577 controls (0.25%) (Fisher's exact test, 2-sided P = 0.15). Patient 5237_3 carried a large deletion (235.2 kb) of the full gene, transmitted by the mother. Patient AU038_3 carried a smaller deletion of 49.8 kb of the first two exons, transmitted by the mother. Both deletions overlap the segmental duplications of BP4 and have been reported to accompany the majority of microduplications involving *CHRNA7* [51]. However, in patient 5237_3, the two CNVs are present on distinct parental chromosomes since the *CHRNA7* duplication and the *ARHGAP11B* deletion are paternally and maternally inherited, respectively. Finally, the third patient, 6319_3, carried a paternally-inherited BP1-BP2 deletion of 468 kb, removing

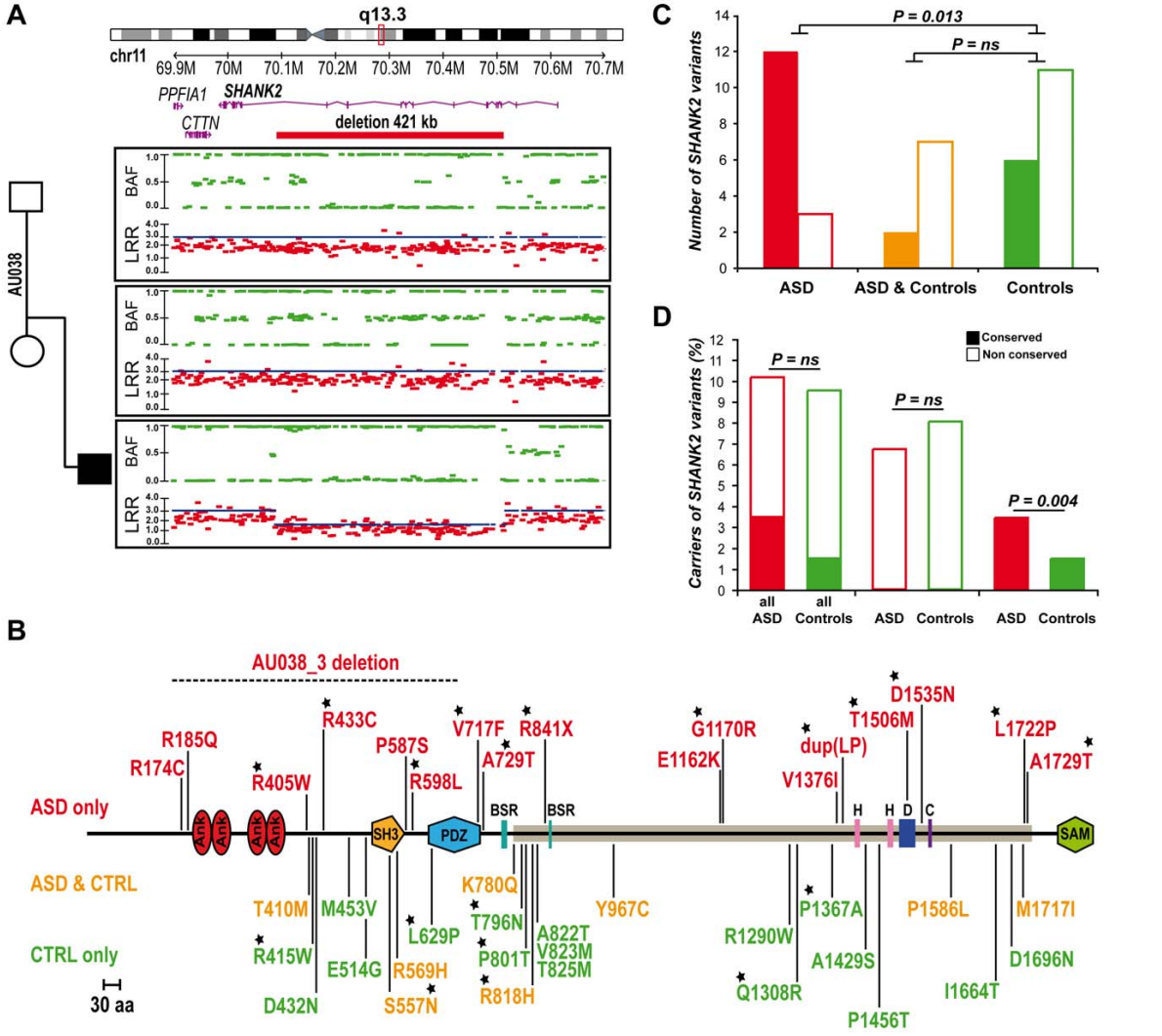


Figure 2. *SHANK2* mutations in patients with ASD. A. A heterozygous deletion of *SHANK2* was identified with the Illumina Human 1M-Duo SNP array in a patient with autism (AU038_3). The deletion spans 421 kb on chromosome 11q13.3, covers twelve exons of the human *SHANK2* and is not present in the parents. Each dot shows Log R Ratio (LRR; in red) and B allele frequency (BAF; in green). QuantiSNP score is represented with a blue line and indicates the deletion size. B. Location of the CNV and sequence variants (from this study and Berkel *et al.* 2010) along the SHANK2 protein: in red the variations specific to ASD, in orange the variations shared by ASD and controls and in green the variations specific to controls [26]. The breakpoints of the *SHANK2* deletion in AU038_3 are represented with a dotted line on the protein. Stars indicate the variants affecting conserved amino acids. C. A total of 40 variants were identified and variants affecting conserved amino acids in other SHANK proteins are enriched in patients with ASD ($n_{conserved} = 12$ and $n_{non\text{-}conserved} = 3$) compared with controls ($n_{conserved} = 6$ and $n_{non\text{-}conserved} = 11$) (Fisher's exact test 1-sided, $P = 0.013$, $OR = 6.83$, 95% $IC = 1.19–53.40$). D. The percentage of carriers of *SHANK2* variants in patients with ASD and Controls. Variants affecting a conserved amino acid among the SHANK proteins are enriched in patients with ASD ($n_{conserved} = 29$ and $n_{non\text{-}conserved} = 822$) compared with controls ($n_{conserved} = 16$ and $n_{non\text{-}conserved} = 1074$) (Fisher's exact test 1-sided, $P = 0.004$, $OR = 2.37$, 95% $CI = 1.23–4.70$). Open squares and filled squares represent the non-conserved and conserved amino acids, respectively. ANK, Ankyrin repeat domain; SH3, Src homology 3 domain; PDZ, postsynaptic density 95/Discs large/zona occludens-1 homology domain; SAM, sterile alpha motif domain; BSR, brain specific region; H, homer binding site; D, dynamin binding site; C, cortactin binding site. The proline-rich region is represented as a horizontal gray line.
doi:10.1371/journal.pgen.1002521.g002

NIPA1, *NIPA2*, *CYFIP1*, and *TUBGCP5*. This deletion was observed in 4 of 1257 patients with ASD (0.32%) and in 4 of 1577 controls (0.25%) (Fisher's exact test, 2-sided $P = 0.74$). The BP1-BP2 deletion is associated with phenotypic variability and has been reported in individuals with neurodevelopmental disorders [20], schizophrenia [53,60], ASD [44–46,49], and epilepsy [61]. In a recent screen for large CNVs (>400 kb) performed on 15,767 children with ID and various congenital defects, and 8,329 unaffected adult controls [20], deletions affecting *CYFIP1*, *NIPA1*, *NIPA2* and *TUBGCP5* were associated with neurodevelopmental disorder ($P = 4.73 \times 10^{-6}$), epilepsy ($P = 1.48 \times 10^{-3}$) and autism ($P = 1.99 \times 10^{-2}$).

Table 1. *ProSAP1A/SHANK2* variations identified in 851 patients with ASD and 1,090 controls.

	Detected variants			Conservation in SHANK proteins[d]	Frequency		Study
	Exon	Nucleotide/dbSNP[a]	Amino acid		ASD (n = 851)	Controls (n = 1090)	
ASD only	E11	G70344397A	R405W	Yes (S1 & S3)	1	0	Berkel *et al.* 2010
	E11	G70344284A	R443C	Yes (S3)	1	0	This study
	E13	G70322214A	P587S	No	1	0	Berkel *et al.* 2010
	E14	C70222501A	R598L	Yes (S3)	1	0	This study
	E17	C70026597A	V717F[b]	Yes (S1 & S3)	1	0	This study
	E17	C70026561T	A729T[b]	Yes (S3)	1	0	This study
	E22	G70014059A	R841X	Yes (S3)	1	0	Berkel *et al.* 2010
	E24	C70010562T	E1162K	No	1	0	This study
	E24	C70010538T	G1170R[b]	Yes (S1 & S3)	1	0	This study
	E24	C70009920T	V1376I	No	1	0	This study
	E24	dup(AACGGT) 70009882–70009887	dup(LP) 1387–1388	Yes (S1 & S3; S1)	1	0	Berkel *et al.* 2010
	E24	G70009529A	T1506M	Yes (S1 & S3)	1	0	Berkel *et al.* 2010
	E24	C70009443T	D1535N[b]	Yes (S1 & S3)	1	0	This study
	E25	A69997007G	L1722P[b]	Yes (S1 & S3)	1	0	This study
	E25	C69996987T	A1729T	Yes (S3)	1	0	Berkel *et al.* 2010
ASD & Controls	E11	G70344381A	T410M	No	1	2	This study & Berkel *et al.* 2010
	E13	C70322303T	S557N	Yes (S1 & S3)	9	3	This study & Berkel *et al.* 2010
	E13	C70322267T	R569H	No	17	28	This study & Berkel *et al.* 2010
	E21	T70016189G rs55968949	K780Q	No	4	4	This study & Berkel *et al.* 2010
	E22	C70014127T rs117843717	R818H[b,c]	Yes (S3)	8	7	This study & Berkel *et al.* 2010
	E24	T70011146C rs62622853	Y967C	No	27	38	This study & Berkel *et al.* 2010
	E24	G70009289A	P1586L	No	4	1	This study & Berkel *et al.* 2010
	E25	C69997021T	M1717I	No	2	2	This study & Berkel *et al.* 2010
Controls only	E11	G70344367A	R415W	Yes (S3)	0	1	Berkel *et al.* 2010
	E11	C70344316T	D432N	No	0	1	Berkel *et al.* 2010
	E11	T70344253C	M453V	No	0	1	This study
	E12	T70330877C	E514G	No	0	1	Berkel *et al.* 2010
	E15	A70185399G	L629P	Yes (S1 & S3)	0	1	This study
	E21	G70016140T	T796N	Yes (S3)	0	1	Berkel *et al.* 2010
	E21	G70016126T	P801T	Yes (S1 & S3)	0	1	Berkel *et al.* 2010
	E22	C70014116T	A822T	No	0	1	This study
	E22	C70014113T	V823M	No	0	1	This study
	E22	G70014106A	T825M	No	0	1	Berkel *et al.* 2010
	E24	G70010178A	R1290W	No	0	1	This study
	E24	T70010123C	Q1308R	Yes (S3)	0	1	This study
	E24	G70009947C	P1367A	Yes (S1 & S3)	0	1	This study
	E24	A70009759G	A1429S	No	0	2	Berkel *et al.* 2010
	E24	G70009680T	P1456T	No	0	1	Berkel *et al.* 2010
	E25	A69997181G	I1664T	No	0	2	Berkel *et al.* 2010
	E25	C69997086T	D1696N	No	0	1	Berkel *et al.* 2010
Total variants					87 (10.2%)	104 (9.5%)	Fisher's exact test 1-sided, P = 0.34, OR = 1.08, 95% CI = 0.79–1.47
Total variants with MAF<1%					43 (5.1%)	38 (3.5%)	Fisher's exact test 1-sided, P = 0.06, OR = 1.47, 95% CI = 0.92–2.37
Total conserved variants					29 (3.4%)	16 (1.5%)	Fisher's exact test 1-sided, P = 0.004, OR = 2.37, 95% CI = 1.23–4.70

Table 1. Cont.

[a]Nucleotide positions are according to *NM 012309.3* from NCBI36/hg18 on the positive DNA strand; The patients with ASD and the controls used for this analysis came from this study (455 ASD & 431 controls) and from the study of Berkel *et al.* 2010 (396 ASD & 659 controls);
[b]A screening of V717F, A729T, R818H, G1170R, D1535N and L1722P was performed in 948 subjects from the Human Genome Diversity Panel (V717F = 0/948; A729T = 0/948; R818H = 5/948; G1170R = 0/948; D1535N = 0/948; L1722P = 0/948);
[c]A screen of R818H was performed in additional patients and controls (ASD 32/3250 (1.0%); controls 27/2030 (1.33%); Fisher's exact test 2-sided, P = 0.28). Fisher's exact test was used for statistical analysis;
[d]"Yes" indicates if amino acid is conserved in SHANK1 (S1), SHANK3 (S3) or both (S1 & S3); MAF, Minor Allele Frequency.
doi:10.1371/journal.pgen.1002521.t001

Several additional CNVs also altered compelling candidate genes for susceptibility to ASD. In patient AU038_3 we detected a previously unreported paternally inherited intronic duplication of *CAMSAP1L* on chromosome 1q32.1, coding for a calmodulin regulated spectrin-associated protein highly expressed in the brain. Patient 5237_3 carried a *de novo* deletion altering the coding sequence of the tyrosine phosphatase *DUSP22* on chromosome 6p25.3 and a maternally inherited intronic duplication of *NLGN1* on chromosome 3q26.3 [9]. These CNVs were observed at similar frequencies in patients with ASD compared with controls. *DUSP22* deletions were observed in 8 of 1257 patients with ASD (0.64%) and in 14 of 1577 controls (0.89%), while *NLGN1* intronic duplications were observed in 60 of 1257 patients with ASD (4.77%) and in 62 of 1577 controls (3.93%). Finally, patient 6319_3 carried an unreported maternally inherited intronic deletion of contactin *CNTN4*, a gene on chromosome 3p26.3 associated with ASD [62], as well as a paternally inherited deletion within the protocadherin *PCDHA1-10* gene cluster on chromosome 5q31.3. Interestingly, this deletion removes the first exon of both *PCDH8* and *PCDH9* and was significantly less frequent in patients with ASD compared with controls (ASD: 62 of 1257; controls: 132 of 1577; Fisher's exact test, 2-sided P = 0.0003; OR = 0.57; 95% CI = 0.41–0.78).

We also analyzed the genome of the Swedish control SWE_Q56_508 carrying the *SHANK2* splice mutation using the Human Omni2.5 BeadChip array from Illumina (Figure 3B). Two close duplications on 2p25.3 were detected, altering four genes, *LOC391343*, *SNTG2*, *PXDN* and *MYT1L*. The inheritance of these two duplications could not be investigated, because DNA samples from the parents were not available. However, 2 of 1577 controls also carried of the same close duplications, suggesting that these CNVs are located on the same chromosome. Among the affected genes, syntrophin-γ2 (*SNTG2*) and myelin transcription factor 1-like (*MYT1L*) are expressed in the brain. Alterations of *SNTG2* and *MYT1L* have been previously reported in patients with ASD [20,63,64] and schizophrenia [65], respectively. SNTG2 is a scaffolding protein interacting with the NLGN3/4X proteins [66]

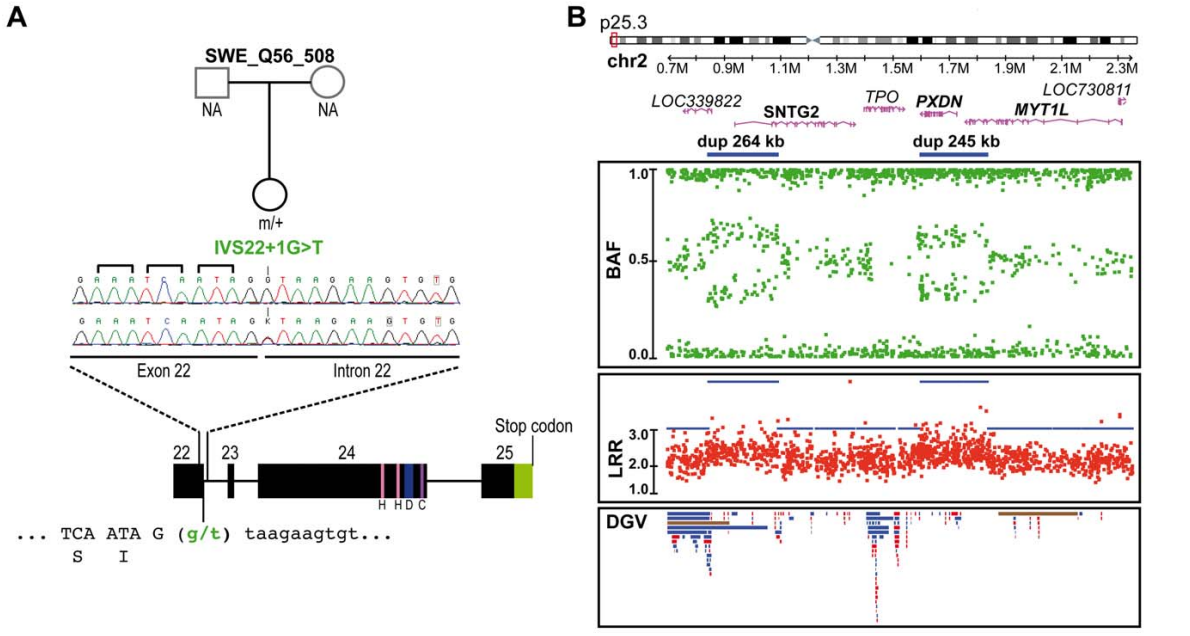


Figure 3. Genetic alterations identified in the control subject SWE_Q56_508. A. *SHANK2* splice mutation (IVS22+1G>T) detected in a Swedish female control, SWE_Q56_508. The mutation altered the donor splicing site of exon 22 and led to a premature stop in all *SHANK2* isoforms except for the *AF1411901* isoform, where it altered the protein sequence (G263V). B. CNVs in the same individual altering *LOC339822*, *SNTG2*, *PXDN* and *MYT1L*. The two close duplications span 264 kb and 245 kb on chromosome 2 and altered *LOC339822* and *SNTG2*, and *PXDN* and *MYT1L*, respectively. Dots show the B allele frequency (BAF; in green), Log R ratio (LRR; in red), and QuantiSNP score (in blue). Lower panel: all CNVs listed in the Database of Genomic Variants (DGV) are represented: loss (in red), gain (in blue), gain or loss (in brown). H, homer binding site; D, dynamin binding site; C, cortactin binding site.
doi:10.1371/journal.pgen.1002521.g003

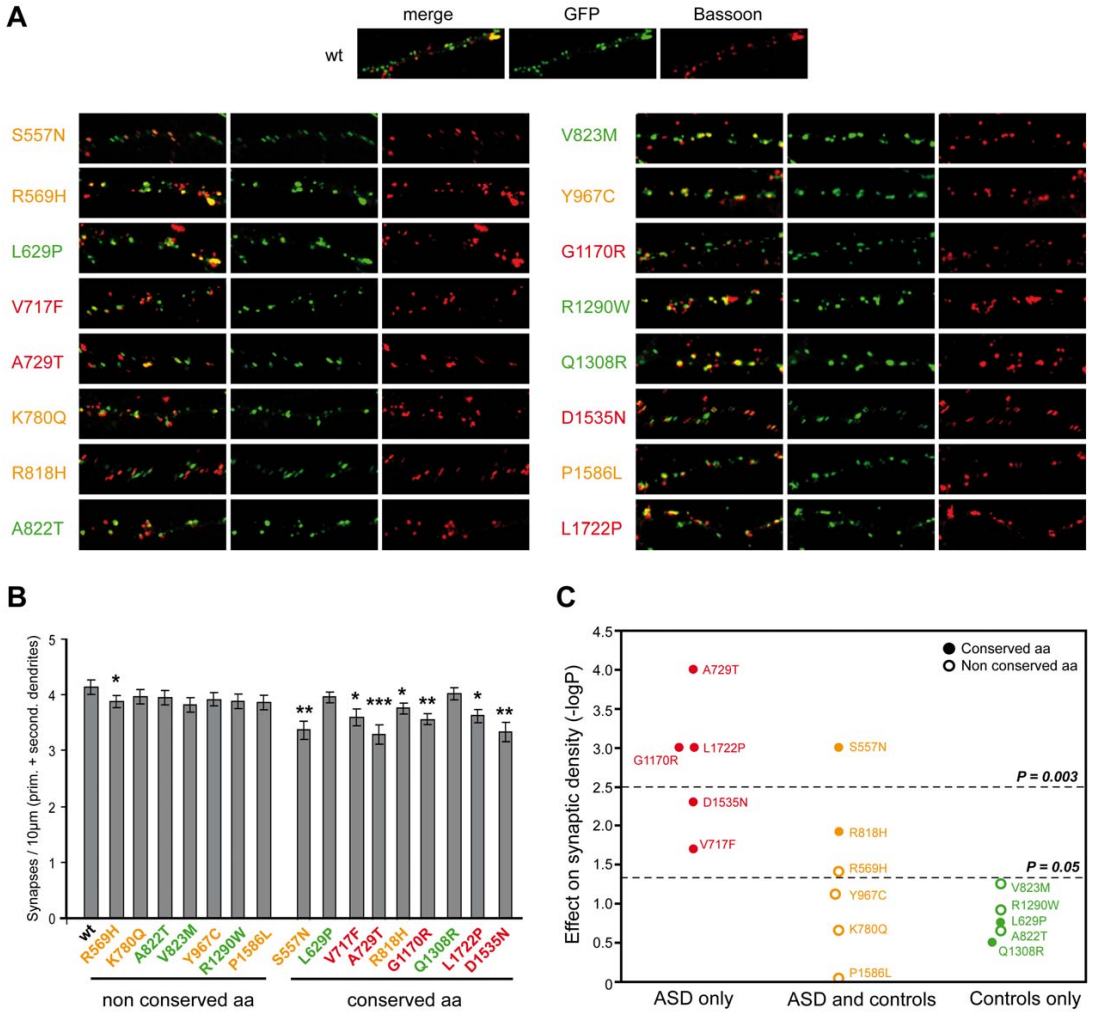


Figure 4. Characterization of the functional impact of *SHANK2* mutations in cultured neuronal cells. A. The colocalization of *ProSAP1A/SHANK2*-EGFP (postsynaptic marker) and Bassoon (presynaptic marker) indicated that the mutations did not disturb the formation of SHANK2 clusters at excitatory synapses along the dendrites. B. The quantification of synapse density was performed on 20 transfected hippocampal neurons per construct from at least three independent experiments. The majority of the *ProSAP1A* variants affecting a conserved amino acid among SHANK proteins reduced significantly the synaptic density compared with the variants that affect amino acid non conserved among SHANK proteins (Mann-Whitney U-test: $n_{WT} = 20$, $n_{mut} = 20$; $U_{S557N} = 82.5$, $p_{S557N} = 0.001$; $U_{R569H} = 124$, $p_{R569H} = 0.04$; $U_{L629P} = 149$, $p_{L629P} = 0.17$; $U_{V717F} = 114$, $p_{V717F} = 0.02$; $U_{A729T} = 73$, $p_{A729T} = 0.000$; $U_{K780Q} = 154$, $p_{K780Q} = 0.221$; $U_{R818H} = 108$, $p_{R818H} = 0.012$; $U_{A822T} = 154.5$, $p_{A822T} = 0.224$; $U_{V823M} = 129$, $p_{V823M} = 0.056$; $U_{Y967C} = 134$, $p_{Y967C} = 0.076$; $U_{G1170R} = 78$, $p_{G1170R} = 0.001$; $U_{R1290W} = 142$, $p_{R1290W} = 0.121$; $U_{Q1308R} = 162$, $p_{Q1308R} = 0.314$; $U_{D1535N} = 97$, $p_{D1535N} = 0.005$; $U_{P1586L} = 137$, $p_{P1586L} = 0.910$; $U_{L1722P} = 79$, $p_{L1722P} = 0.001$, *$p<0.05$, **$p<0.01$, ***$p<0.001$). **C.** Effect of the variants on synaptic density. The y-axis represents $-\log$ P compared to WT (P obtained with Mann-Whitney test). After Bonferroni correction for 16 tests, only P values<0.003 were considered as significant. Variants represented in red were specific to ASD, in orange were shared by ASD and controls, and in green were specific to the controls. Open circles and filled circles represent non conserved and conserved amino acids, respectively. Prim, primary; second, secondary.
doi:10.1371/journal.pgen.1002521.g004

and a component of the dystrophin glycoprotein complex [67]. MYT1L is a myelin transcription factor required to convert mouse embryonic and postnatal fibroblasts into functional neurons [68].

Discussion

Deleterious *SHANK2* variations are enriched in patients with ASD, but also observed in controls

The identification of mutations in synaptic proteins such as NRXN1, NLGN3/4X and SHANK2/3 has demonstrated that a synaptic defect might be at the origin of ASD [5,6]. Here we confirm the presence of *SHANK2 de novo* deletions in individuals with ASD, with a prevalence of 0.38% (1/260) in our cohort of ASD patients analyzed with the Illumina 1M SNP array. This frequency is similar to the one reported previously by the AGP in a larger sample of 996 patients with ASD (0.2%) [9]. *SHANK2* deletions altering exons were not detected in controls, in agreement with previous findings [9,26]. As reported for *SHANK3* [7], no other coding variations were detected in the remaining *SHANK2* allele of the deletion carriers, suggesting that, in some

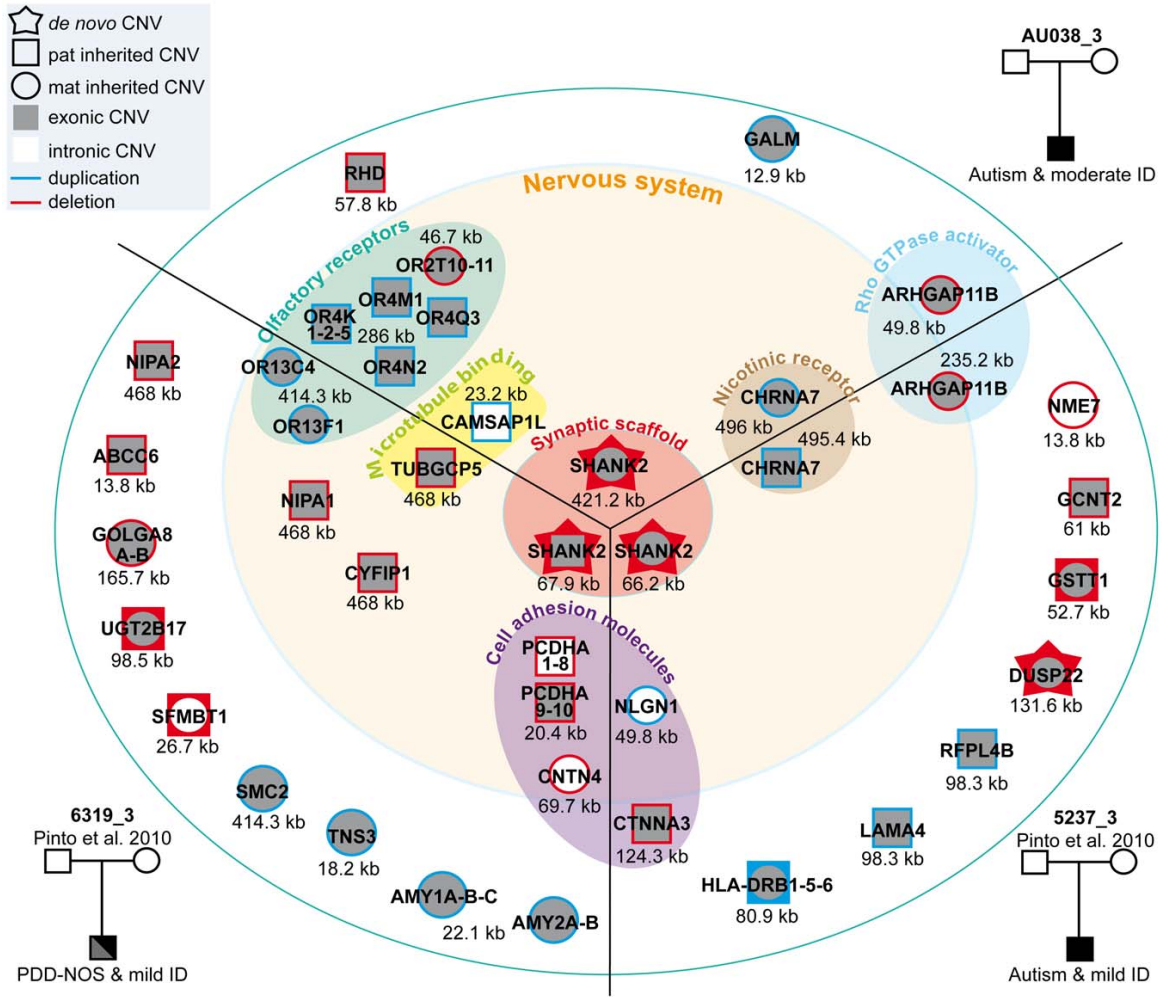


Figure 5. Characterization of CNVs in three patients carrying a *de novo* deletion of *SHANK2*. Paternally or maternally inherited CNVs are indicated by squares and circles, respectively. *De novo* CNVs are indicated by stars. Deletions and duplications are indicated in red and blue, respectively. CNVs hitting exons or only introns are filled with grey and white, respectively. Squares and circles within star represent *de novo* CNV of paternal or maternal origin; circles within squares represent CNV inherited by father or mother. ABCC6, ATP-binding cassette, sub-family C, member 6 pseudogene 2; ADAM, ADAM metallopeptidase; AMY1, amylase (salivary); AMY2A, amylase (pancreatic); ARHGAP11B, Rho GTPase activating protein 11B; CAMSAP1L1, calmodulin regulated spectrin-associated protein 1-like 1; CHRNA7, cholinergic receptor, nicotinic, alpha 7; CNTN4, contactin 4; CTNNA3, catenin (cadherin-associated protein), alpha 3; CYFIP1, cytoplasmic FMR1 interacting protein 1; DUSP22, dual specificity phosphatase 22; GALM, galactose mutarotase; GCNT2, glucosaminyl (N-acetyl) transferase 2; GOLGA, golgi autoantigen, golgin subfamily a; GSTT1, glutathione S-transferase theta 1; HLA-DRB, major histocompatibility complex, class II, DR beta; LAMA4, laminin, alpha 4; NIPA, non imprinted in Prader-Willi/Angelman syndrome; NLGN1, neuroligin 1; NME7, non-metastatic cells 7; OR, olfactory receptor; PCDHA, protocadherin alpha; RFPL4B, ret finger protein-like 4B; RHD, Rh blood group, D antigen; SFMBT1, Scm-like with four mbt domains 1; SHANK2, SH3 and multiple ankyrin repeat domains 2; SMC2, structural maintenance of chromosomes 2; TNS3, tensin 3; TUBGCP5, tubulin, gamma complex associated protein 5; UGT2B17, UDP glucuronosyltransferase 2 family, polypeptide B17.
doi:10.1371/journal.pgen.1002521.g005

individuals, a *de novo* deletion of a single allele of *SHANK2* might be sufficient to increase the risk for ASD. In one case, a patient carried two rare *SHANK2* variants predicted as deleterious and inherited from different parents, indicating that they were separate alleles.

For the remaining *SHANK2* variants, patients were heterozygous for non-synonymous rare variations inherited from one of their parents (Figure S3). Since parents were apparently asymptomatic, the causative role of these variants in ASD remains difficult to ascertain. However, we observed a significant enrichment of *SHANK2* variants affecting conserved amino acids in patients with ASD compared with controls. This was also the case in the previous mutation screening by Berkel *et al.* 2010 [26]. The majority of the variants affecting conserved residues and identified in the patients were shown to alter the ability of *SHANK2* to increase the number of synapses *in vitro*. Importantly, the assays performed in this study show that the variants can potentially impact on the function of the protein, but they do not confirm that they have deleterious effects on neuronal function *in vivo* in people that carry them. However, these results are consistent with

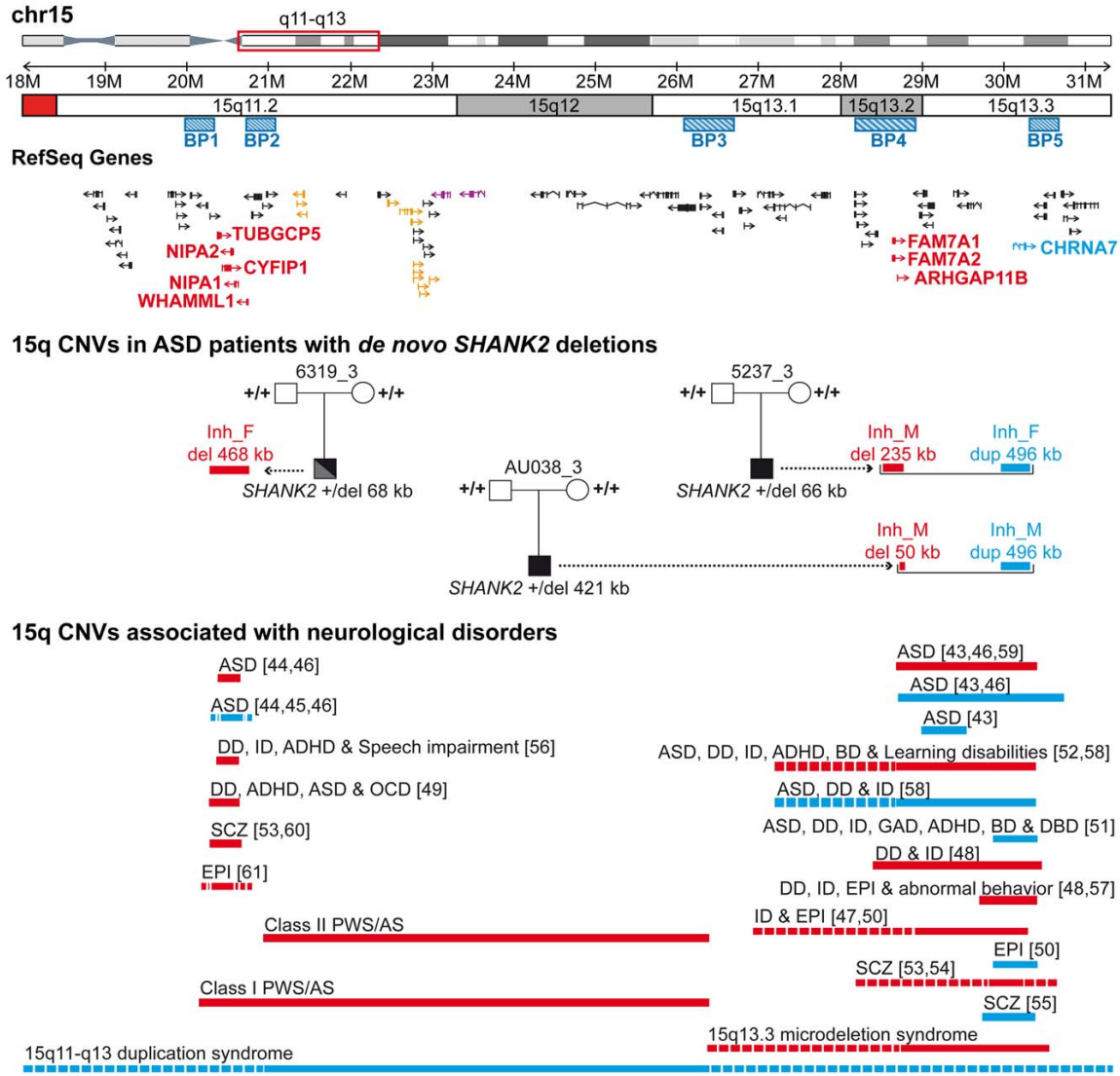


Figure 6. Inherited 15q11–q13 CNVs identified in three ASD patients carrier of a *de novo* SHANK2 deletion. Deletions (del) and duplications (dup) are indicated in red and blue, respectively. Paternally and maternally imprinted genes are indicated in yellow and pink, respectively. Genes altered by the CNVs are indicated in blue or red. The bottom part of the figure indicates the location of the deletions/duplications previously associated with neuropsychiatric disorders [43–61]. BP, breakpoint; Inh_M, inherited by mother; Inh_F, inherited by father; AS, Angelman syndrome; ASD, Autism spectrum disorders; ADHD, attention deficit-hyperactivity disorder; BP, bipolar disorder; DD: developmental delay; DBD, disruptive behavior disorder; EPI, epilepsy; GAD, generalized anxiety disorder; OCD, obsessive-compulsive disorder; ID, intellectual disability; PWS, Prader-Willi syndrome; SCZ, schizophrenia.
doi:10.1371/journal.pgen.1002521.g006

previous findings showing that inherited variants of SHANK2 and SHANK3 cause synaptic defects *in vitro* [7,69,70]. Recently, Berkel *et al.* 2011 showed that two inherited (L1008_P1009dup, T1127M) and one *de novo* (R462X) SHANK2 mutations identified in patients with ASD affect spine volume and reduced Shank2 cluster sizes [70]. This deleterious effect was also observed *in vivo* since mice expressing rAAV-transduced Shank2-R462X present a specific long-lasting reduction in miniature postsynaptic AMPA receptor currents [70].

In patients, the only feature associated with carriers of *SHANK2* mutations compared with other patients was a trend for low IQ ($P = 0.025$, $OR = 3.75$, 95% $CI = 1.1–20.0$) (Table S8). But, as observed for *SHANK3* mutations, this correlation could differ from one individual to another (i.e. the patient with a *SHANK2 de novo* stop mutation reported by Berkel *et al.* 2010 presented with high-functioning autism [26]).

Our result also showed that potentially deleterious *SHANK2* variants were detected in a heterozygous state in parents and in the general population without causing severe phenotypic consequences. Indeed, we showed that almost 5% of the Finnish population is heterozygous for the *SHANK2* R818H variation, which modifies a conserved amino acid and is associated with

lower synaptic density *in vitro*. Furthermore, we identified a *SHANK2* splice site mutation in a control female without any apparent psychiatric disorders. Similarly, two frame-shift mutations and one splice site mutation of *SHANK2* are listed in dbSNP and in the 1000 genomes project [71]. These nonsense variations should be interpreted with caution since none of them has been validated by Sanger sequence technology. Taken together, variants affecting conserved amino acids of SHANK2 might act as susceptibility variants for ASD, but, in some cases, additional genetic, epigenetic or environmental factors seem to be necessary for the emergence of the disorder.

Additional CNVs in subjects with *SHANK2* mutations may modulate the risk for ASD

In order to detect risk and protective genetic factors, we analyzed the CNV burden of the individuals carrying deleterious variations of *SHANK2*. Notably, the three ASD patients with *de novo SHANK2* deletions also carried CNVs on chromosome 15q11–q13, a region associated with ASD [43,47,48,50–52,72]. In contrast, the patient reported by Berkel *et al.* 2010, who did not meet all the diagnostic criteria for ASD, seemed to have no CNV at chromosome 15q [26]. Although the probability to observe the co-occurrence of a *de novo SHANK2* deletion and a duplication of *CHRNA7* at 15q is very low, two of the three patients carrying a *de novo SHANK2* deletion also carried the *CHRNA7* duplication. While the numbers are small, this finding could suggest epistasis between these two loci. The role of *CHRNA7* in ASD was recently supported by the observation of low levels of *CHRNA7* mRNA in the post-mortem brain from patients with ASD [73]. Interestingly, it was also found that, in contrast to the gene copy number, the transcript levels *CHRNA7* were reduced in neuronal cells [74] or brain samples with maternal 15q duplication [75]. Finally, functional studies have shown that NLGN and NRXN, which belong to the same synaptic pathway, are key organizers of the clustering of nicotinic receptors at the synapse [76–78]. Therefore the co-occurrence of a deletion of *SHANK2* and a duplication of the nicotinic receptor *CHRNA7* could act together within the same pathway to increase the risk of ASD in patients AU038_3 and 5237_3. In patient 6319_3 carrying the BP1–BP2 deletion, several genes might also play a role in the susceptibility to ASD. Among them, *NIPA1* and *TUBGCP5* encoding a magnesium transporter and a tubulin gamma associated protein, respectively, are highly expressed in the brain. However, the most compelling candidate in the deleted region is *CYFIP1* [45,53], which codes for a binding partner of FMRP, the protein responsible for fragile X syndrome. Both CYFIP1 and FMRP are involved in the repression of synaptic translation [79], one of the major biological mechanisms associated with ASD [80]. Therefore, the co-occurrence of a loss of one copy of *SHANK2* and *CYFIP1* might increase the risk of abnormal synaptic function in patient 6319_3.

If some individuals have a higher risk to develop ASD when a deleterious *SHANK2* variant is present, others individuals may experience a protective effect by additional genetic factors. For example, control SWE_Q56_508 carried a *SHANK2* splice mutation, but clinical examination revealed no major disorders. In addition, this control individual also carried a partial duplication of *SNTG2* and *MYT1L*. Based on a single control subject, it is not possible to formally prove that these additional hits at *SNTG2* and/or *MYT1L* acted as suppressor mutations, counteracting the phenotypic effects of the *SHANK2* splice mutation. However, the encoded proteins may interact with the NRXN-NLGN-SHANK pathway. Both SNTG2 and SHANK2 are scaffolding proteins localized in actin rich structures [81–83] and bind directly to neuroligins [66]. Furthermore, mutations of *NLGN3/4X* identified in patients with ASD decrease their protein binding to SNTG2 [66]. In addition, MYT1L is a myelin transcription factor that is sufficient, with only two other transcription factors, ASCL1 and BRN2, to convert mouse embryonic and postnatal fibroblasts into functional neurons *in vitro* [67]. Therefore, alterations of SNTG2 and/or MYTL1 might modulate synapse physiology and counteract the effect of the *SHANK2* splice site mutation. We recently highlighted the key role of synaptic gene dosage in ASD and the possibility that a protein imbalance at the synapse could alter synaptic homeostasis [6]. In the future, animal models should be developed to test whether the effect of a primary mutation in a synaptic protein complex (e.g. *Shank2*) can be reduced or suppressed by a second mutation (e.g. *Sntg2* or *Myt1l*). A similar suppressor effect has been demonstrated by the decrease of abnormal behavior of the *Fmr1* mutant mice carrying a heterozygous mutation of the metabotropic glutamate receptor mGluR5 [84].

Conclusions and perspectives

In summary, we confirmed that *de novo SHANK2* deletions are present in patients with ASD and showed that several *SHANK2* variants reduce the number of synapses *in vitro*. The genomic profile of the patients carrying deleterious *de novo SHANK2* deletions also points to a possible genetic epistasis between the NRXN-NLGN-SHANK pathway and 15q11–q13 CNVs. *CHRNA7* and *CYFIP1* were already proposed as susceptibility genes for neuropsychiatric disorders [43,45,49,51], and our study provides additional support for this association. Therefore, as previously observed for ID [85], our results suggest that the co-occurrence of *de novo* mutations, together with inherited variations might play a role in the genetic susceptibility to ASD. Finally, our analyses suggest the interesting possibility that deleterious mutations of neuronal genes (e.g. *SNTG2* and *MYT1L*) could potentially counteract the effect of synaptic deleterious mutations (e.g. *SHANK2*). The identification of risk and protective alleles within the same subject is one of the main challenges for understanding the inheritance of ASD. Initial results from the 1000 genomes project has estimated that, on average, each person carries approximately 250 to 300 loss-of-function variants in annotated genes and 50 to 100 variants previously implicated in inherited disorders [71]. To date, it is not clear how many loci can regulate synaptic homeostasis and how these variants interact with each other to modulate the risk for ASD [6]. A better knowledge of these genetic interactions will be necessary to understand the complex inheritance pattern of ASD.

Materials and Methods

Ethics statement

This study was approved by the local Institutional Review Board (IRB) and written inform consents were obtained from all participants of the study. The local IRB are the "Comité de Protection des Personnes" (Île-de-France Hôpital Pitié-Salpêtrière Paris) for France; the Sahlgrenska Academy Ethics committee, University of Gothenburg for Sweden; the local IRB of the medical faculty of JW Goethe University Frankfurt/Main for Germany; the Committee #3 of the Helsinki University Hospital, Finland; the "Comitato Etico IRCCS Fondazione Stella Maris" at Stella Maris Institute, Calambrone (Pisa), Italy; the "Comitato Etico Azienda Ospedaliera-Universitaria Policlinico-Vittorio Emanuele", Catania, Italy.

Patients

Patients with ASD and analyzed for CNV analysis and/or mutation screening are presented in Table S1. Patients were

recruited by the PARIS (Paris Autism Research International Sibpair) study at specialized clinical centers disposed in France, Sweden, Germany, Finland, UK. The Autism Diagnostic Interview-Revised (ADI-R) and Autism Diagnostic Observation Schedule (ADOS) were used for clinical evaluation and diagnosis. In Sweden, in some cases, the Diagnostic Interview for Social and Communication Disorders (DISCO-10) was applied instead of the ADI-R. Patients were included after a clinical and medical check-up with psychiatric and neuropsychological examination, standard karyotyping, fragile-X testing and brain imaging and EEG whenever possible. All patients were from Caucasian ancestry.

The patient AU038_3 with a *de novo SHANK2* deletion is an 11.05 year-old boy diagnosed with autism and moderate ID (Table S2). He was the only child of non-consanguineous parents from European descent. His parents had no relevant personal and familial history of psychiatric or medical illness. He was born at 40 weeks of gestation, after normal pregnancy and delivery. Birth weight, length and occipitofrontal head circumference were 2500 g (5th percentile), 48 cm (22nd percentile) and 31 cm (2nd percentile), respectively. Apgar scores were 7 and 10 at 1 and 5 minutes, respectively. In the first year of life, the pediatrician reports did not mention signs of hypotonia. At 2 months, he was operated for an inguinal hernia. Motor acquisition was apparently normal (sitting at 6 months), but with a late acquisition of walking, at 18 months. Speech was severely delayed, without any apparent regressive phase. Only a few words and sentences appeared when he was 4 y and 6.5 y, respectively. His expressive language remained limited to restrictive sentences, mainly dyssyntaxic. A formal diagnostic assessment for autism was performed when he was 11 years old. The scores of the Autism Diagnostic Interview-Revised (ADI-R) domains were: social 24, communication 23, and behavior 6 (cut-offs for autism diagnosis are 10, 8 (verbal autism) and 3, respectively); the age at first symptoms was before 36 months. Cognitive evaluation with the Kaufman Assessment Battery for Children (K-ABC) showed moderate intellectual deficit (composite score 40). He required assistance with basic activities such as eating and dressing. At examination, he had a normal facial appearance, with a prominent chin. General and neurological examinations were normal, except for hypermetropia and astigmatism. High-resolution karyotype, fragile X testing, MLPA analysis of telomeres and microdeletion/microduplication syndromes, and metabolic screening for inherited disorders of metabolism (urine amino acids, mucopolysaccharides and organic acids, uric acid in blood and urine) were all normal. No significant epileptic event was reported on the electroencephalogram.

The two male patients with *de novo SHANK2* deletions reported by Pinto *et al.* 2010 [9] (5237_3 and 6319_3) shared several clinical features with patient AU038_3. Patient 5237_3 is a Canadian subject diagnosed with autism (based on ADI-R and ADOS) associated with below average non verbal IQ ($<1^{st}$ percentile) and language ($<1^{st}$ percentile). He had minor dysmorphic features including 5^{th} finger clinodactyly and several curled toes, and no history of epilepsy. Patient 6319_3 was recruited in the same geographic area as patient AU038_3 (Grenoble, France) and was clinically diagnosed with PDD-NOS. The ADI-R scores were: social 14, communication 8, behaviors 2 (cut-off for autism: 3); with an age at first symptoms <36 months). He had mild ID as evaluated with the WISC-III (full scale IQ 60, performance IQ 60, verbal IQ 67). His language was delayed (first words 24 m, first sentences 48 m), but functional. He had no history of regression or epilepsy. The physical exam was normal, except for large and prominent ears and flat feet; the neurological exam was also normal. Similarly to patient AU038_3, he had hypermetropia.

The control female carrying the splice site mutation (IVS22+1G>T) was part of a cohort of 172 females recruited for a study on obesity, anthropometrics, and cardiovascular risk factors [37]. In addition, these women were assessed for axis I psychiatric disorders and for personality traits using the Temperament and Character Inventory (TCI) [38] and the Karolinska Scales of Personality (KSP) [39]. This subject had no psychiatric disorders and her TCI and KSP scores were similar to those found in the general population.

Genomic structure and transcripts analysis of *SHANK2*

To define the genomic structure of the human *SHANK2* gene, we used the two reference sequence genes from UCSC (*NM_012309* and *NM_133266*), one human mRNA from GenBank (*DQ152234*) and three Rattus reference sequence genes from UCSC (*NM_201350*, *NM_133441* and *NM_133440*). *SHANK2* is transcribed in four isoforms described in GenBank (*AB208025*, *AB208026*, *AB208027* and *AF141901*) and is composed of 25 exons. Transcript analysis of *SHANK2* was performed in human brain regions from four independent controls (two females and two males) and in human tissues (heart, brain, placenta, lung, liver, skeletal muscle, kidney, pancreas and B lymphoblastoid cell lines) using the Clontech Multiple Tissue cDNA panel (Clontech). Total RNA was isolated from control human brain tissues by the acid guanidinium thiocyanate phenol chloroform method and reverse transcribed by oligodT priming using SuperScript II Reverse Transcriptase (Invitrogen). The PCR was performed with HotStar Taq polymerase (Qiagen) and the protocol used was 95°C for 15 min, followed by 40 cycles at 95°C for 30 sec, 55 to 58°C for 30 sec, 72°C for 30 sec to 1 min, with a final cycle at 72°C for 10 min. PCR primers were designed to detect the ANK domain, the SH3 domain, the PDZ domain, and the SAM domain in order to distinguish the four *SHANK2* isoforms and are indicated in Table S11. All RT-PCR products were directly sequenced. The expression of *SHANK2E* isoform was also studied by SYBR-Green real-time PCR approach. The fluorescence was read with the Applied Biosystems 7500 Real-Time PCR System. Each assay was conducted in three replicates. GAPDH was used for the ΔCt calculation and total brain was used as the reference for relative quantification calculation (RQ). The relative RQ of transcripts was calculated as $2^{-\Delta\Delta CT}$ with the magnitude of upper error as $2^{-(\Delta\Delta CT-SEM)}$-$2^{-\Delta\Delta CT}$ and the magnitude of lower error as $2^{-\Delta\Delta CT}$-$2^{-(\Delta\Delta CT+SEM)}$. The primers specific to *SHANK2E* isoform are indicated in Table S11. *In situ* hybridization was performed essentially as described previously [28]. Transcripts encoding the different *ProSAP1/Shank2* cDNAs (*ProSAP1/Shank2* starting with the PDZ domain, *ProSAP1A*, starting with the SH3 domain and *ProSAP1E/Shank2E*, starting with the ankyrin repeats) were detected with isoform specific S35 labeled cDNA antisense oligonucleotides purchased from MWG-Biotech (Ebersberg, Germany) directed against the ATG regions of the different mRNAs. All variants were evaluated for potential pathogenicity using the HumDIV method for rare alleles of PolyPhen2 [41].

CNV detection and validation

DNA was extracted from blood leukocytes or B lymphoblastoid cell lines. The *SHANK2* CNV was detected with the Illumina Human 1M-Duo BeadChip, which interrogates 1 million SNPs distributed over the human genome. For the Swedish control SWE_Q56_508 carrying the *SHANK2* splice mutation we used the Illumina Human Omni2.5 BeadChip array. The genotyping was performed at the Centre National de Génotypage (CNG) and the Institut Pasteur. Only samples that met stringent quality control (QC) criteria were included: call rate ≥99%; high confidence score

log Bayes factor ≥15; standard deviation of the log R ratio (LRR) ≤0.35 and of the B allele frequency (BAF)≤0.13; number of consecutive probes for CNV detection ≥5; CNV size ≥1 kb. When the QC criteria were met, we used two CNV calling algorithms, QuantiSNP [86] and PennCNV [87], and the CNV viewer, SnipPeep (http://snippeep.sourceforge.net/). To obtain high-confidence calls, the CNVs identified by QuantiSNP were validated by visual inspection of the LRR and BAF values. PennCNV was used to confirm inheritance status of the resulting CNV calls. CNVs were validated by quantitative PCR analysis using the Universal Probe Library (UPL) system from Roche. UPL probes were labeled with FAM and the fluorescence was read with the Applied Biosystems 7500 Real-Time PCR System. Each assay was conducted in four replicates for target region probe-set and control region probe-set. Relative levels of region dosage were determined using the comparative CT method assuming that there were two copies of DNA in the control region. The relative copy number for each target region was calculated as $2^{-\Delta\Delta CT}$ with the magnitude of upper error as $2^{-(\Delta\Delta CT-SEM)}$-$2^{-\Delta\Delta CT}$ and the magnitude of lower error as $2^{-\Delta\Delta CT}$-$2^{-(\Delta\Delta CT+SEM)}$. UPL probes and primers are indicated in Table S12. For comparisons between patients and controls, statistical significance for each CNV was assessed using a 2-sided Fisher's exact test.

Mutation screening

The 24 coding exons of *SHANK2* were amplified and sequenced for mutation screening. The PCR was performed on 20–40 ng of genomic DNA template with HotStar Taq polymerase from Qiagen for all exons the protocol used was 95°C for 15 min, followed by 35–40 cycles at 95–97°C for 30 sec, 55–62°C for 30 sec, 72°C for 30 sec to 90 sec, with a final cycle at 72°C for 10 min. Sequence analysis was performed by direct sequencing of the PCR products using a 373A automated DNA sequencer (Applied Biosystems). Genotyping of R185Q, V717F, A729T, R818H, G1170R, D1535N and L1722P was performed by direct sequencing or Taqman SNP Genotyping Assays system from Applied Biosystems designed with Custom TaqMan Assay Design Tool. All primers are indicated in Table S9. Enrichment of *SHANK2* variations in the ASD sample compared with controls was assessed using a 1-sided Fisher's exact test (hypothesizing that cases will show an excess of *SHANK2* variants compared to controls).

In vitro mutagenesis and transfection studies in hippocampal neurons

Rat *GFP-ProSAP1A* (*Shank2A*) cDNA was mutated according to the human mutations using the site directed mutagenesis kit (Stratagene). The mutagenesis primers were listed in Table S10. We have tested all the variants (n = 16) identified in our first screen of 230 patients with ASD and 230 controls: 5 were detected only in patients (V717F, A729T, G1170R, D1535N and L1722P), 6 were detected in patients and controls (S557N, R569H, K780Q, R818H, Y967C and P1586L) and 5 were only found in controls (L629P, A822T, V823M, R1290W and Q1308R). All mutated amino acids were conserved among human, rat and mouse ProSAP1/Shank2. All cDNAs were sequenced and subsequently tested for expression by Western blot analysis. After expression of the constructs in Cos7 cells, the cell homogenate was separated on a gel, transferred to a nitrocellulose membrane and subsequently protein bands were detected using a rabbit anti-GFP antibody. Thereafter, the cDNAs were transfected into primary hippocampal neurons. Cell culture experiments of rat hippocampal primary neurons (embryonic day 18–21: E18-21) were performed as described previously [88]. In brief, after preparation, hippocampal neurons were seeded on poly-l-lysine (0.1 mg/ml; Sigma-Aldrich, Steinheim, Germany) coated coverslips at a density of 4×10^4 cells/well (transfection experiments) or 2×10^4 cells/well (immunological staining). Cells were grown in Neurobasal medium (Invitrogen, Karlsruhe, Germany), complemented with B27 supplement (Invitrogen), 0.5 mM L-glutamine (Invitrogen), and 100 U/ml penicillin/streptomycin (Invitrogen) and maintained at 37°C in 5% CO_2. Hippocampal cells were transfected using Lipofectamine 2000, according to the manufacturer's recommendation (Invitrogen). Fluorescence images were obtained using a camera attached to a fluorescence microscope. For immunofluorescence, the primary cultures were fixed with ice cold 4% paraformaldehyde/1.5% sucrose/PBS for 20 min at 4°C and processed for immunohistochemistry. After washing three times with 1× PBS for 5 min at room temperature the cells were permeabilized for 3 min on ice in a buffer containing 0.1% Triton X-100/0.1% Na-Citrate/PBS and washed again three times with 1× PBS. Blocking was performed with 10% fetal calf serum/PBS for 1 h at room temperature followed by incubation with the primary antibody (mouse anti-Bassoon) overnight at room temperature. After a further washing-step the cells were incubated with the secondary antibody coupled to Alexa555 (red) (Molecular Probes, Invitrogen) for 90 min at room temperature, washed first with 1×PBS and then with aqua bidest for 5 min and mounted in Mowiol (with or without DAPI for staining of the nucleus). All animal experiments were performed in compliance with the guidelines for the welfare of experimental animals issued by the Federal Government of Germany, the National Institutes of Health and the Max Planck Society.

Image acquisition and quantification

In morphological studies, dendrites were considered primary when processes extended directly from the cell body, and secondary when processes branched off primary dendrites. Twenty transfected neurons were chosen randomly for quantification from at least three independent experiments for each construct. Morphometric measurements were performed using Axiovision Zeiss microscope and Axiovision software with a 40× magnification. For the quantification of excitatory synapse number, cells were counterstained with anti-Bassoon antibodies. From randomly chosen transfected neurons, Bassoon-positive spots from primary dendrites were counted and the length of dendrites was measured. The total number of spines was expressed as density per 10 µm length of dendrite. Measured data were exported to Excel software (Microsoft), and the data of each variant were compared by using the Mann-Whitney U test. The comparisons of synaptic density for each phenotypic or conservation categories were performed using the Student's t test.

Supporting Information

Figure S1 *SHANK2E* expression in human multiple tissue panel and in rat embryos. A. Quantitative RT-PCR in human tissues. Primers and probe were designed to detect *SHANK2E* isoform. GAPDH was used for the ΔCt calculation and total brain was used as the reference for relative quantification calculation (RQ ± SEM). B. *In situ* hybridization of rat fetus sagittal sections with *ProSAP1/Shank2* isoform specific oligonucleotides. The *ProSAP1/Shank2* isoform starting with the PDZ domain is solely expressed in brain, brain stem and medulla. The same holds true for the *ProSAP1A/Shank1A* isoform that starts with the SH3 domain. In some sections bone tissue also gave some moderately positive signals. The expression of *ProSAP1E/Shank1E* (with the ankyrin

repeats) is especially seen in the liver and some glandular tissue. In the brain, the *ProSAP1E/Shank2E* mRNA is only detectable within the cerebellum (arrow).
(TIF)

Figure S2 Characterization and validation of the *SHANK2* CNV in family AU038. A. Pedigree of the AU038 family showing that the deletion is *de novo* on the maternal chromosome. SNPs were genotyped using the Illumina 1M duo array. B. *SHANK2* CNV validation by quantitative PCR of exon E4–E6, E15–E17 of *SHANK2* using genomic DNA from the father, mother and the proband of family AU038. Results from QPCR analysis confirmed that the deletion is *de novo* and removes exon E5 to Exon E16. Bars represent mean of RQ ± SEM.
(TIF)

Figure S3 Pedigree of families with ASD carrier of non-synonymous *SHANK2* mutations. Variations specific to ASD or shared by ASD and controls are indicated in red and orange, respectively. Clinical phenotypes are specified in the figure.
(TIF)

Figure S4 Functional analysis of *shank2* mutations in cultured hippocampal neurons. A. Rat hippocampal neurons transfected with wild-type or mutated *ProSAP1A/SHANK2*-EGFP cDNA constructs were detected by EGFP expression. B. The quantification of dendrite number was performed on 20 transfected hippocampal neurons per construct from at least 3 independent experiments. No mutation affected the number of dendrites per neuron (Bars represent mean ± SD, Mann-Whitney U test and Kolmogorov-Smirnov Z test; no significant differences; $n_{WT} = 20$, $n_{mut} = 20$). Western blot analysis of *ProSAP1A/SHANK2-EGFP* cDNA constructs revealed similar sizes of WT fusion *ProSAP1A/SHANK2*-EGPF protein and the constructs carrying the mutations. C. The variants identified in patients with ASD were associated with reduced synaptic density compared with those identified in controls. D. The variants affecting conserved amino acids in other SHANK proteins were associated with reduced synaptic density compared with those affecting non conserved amino acids. Variations represented in red were specific to ASD, those in orange were shared by ASD and controls, and those in green were specific to controls.
(TIF)

Table S1 Description of the population of patients with ASD analyzed in this study.
(DOC)

Table S2 Clinical description of the patients carrying predicted deleterious *SHANK2* variations. Abbreviations: ADI-R, Autism Diagnosis Interview-Revised; ASD, autism spectrum disorder; CT, computed tomography; ERG, electroretinogram; F, female; FSIQ, full scale IQ; HC, head circumference; K-ABC, Kaufman Assessment Battery for Children; M, male; MRI, magnetic resonance imaging; MZ, monozygotic; NA, not available; ND, not done; PDD-NOS, pervasive developmental disorder not otherwise specified; PET, positron emission tomography; PIQ, performance IQ; PEP-R, Psychoeducational Profile-Revised; VIQ, verbal IQ; WISC-III, Wechsler Intelligence Scale for Children-Third Edition; WPPSI, Wechsler Preschool and Primary Scale of Intelligence; IQ, intellectual quotient.
(DOC)

Table S3 Frequency of *SHANK2* R818H variation in 3250 patients with ASD and 2013 controls. OR, odds ratio; P, p-value.
(DOC)

Table S4 Frequency of *SHANK2* R818H variation in 948 individuals from the Human Genome Diversity Panel.
(DOC)

Table S5 Evolutionary conservation of SHANK2 protein sequence. Variations identified only in patients with ASD, only in controls or shared by patients and controls are indicated in red, green and orange, respectively. Hu, human; Ch, chimpanzee; Ma, macaque; Ra, rat; Mo, mouse; Ck, chicken; Zn, zebrafinsh; Zf, zebrafish; Xe, xenopus.
(DOC)

Table S6 List of all CNVs observed in ASD patients carrying a *de novo* deletion of *SHANK2*. The 6319_3 and 5237_3 patients were described by the AGP [9]. QSNP, QuantiSNP; PCNV, PennCNV; IP, iPattern.
(DOC)

Table S7 Distribution of *SHANK2* variants affecting conserved or non conserved amino acids. All variants came from this study (26) and from Berkel *et al.* 2010 (24). *Several variants shared by patients with ASD and controls were identified in both studies.
(DOC)

Table S8 Clinical comparison of patients with ASD carrying SHANK2 variants with the rest of the cohort of patients. We used the Wilcoxon test for the continuous traits and the Fisher's exact test (2-sided) for the discontinuous traits. OR is given with 95% confidence interval. OR, odds ratio; P, p-value; ADI-R, Autism Diagnosis Interview revised.
(DOC)

Table S9 Primers used for mutation screening.
(DOC)

Table S10 Primers used for *in vitro* mutagenesis.
(DOC)

Table S11 Primers used for mRNA analysis of *SHANK2* isoforms. * Primers were used for relative quantification study of *SHANK2E* isoform. The other primers were used for RT-PCR analysis of each *SHANK2* isoform.
(DOC)

Table S12 Primers used for CNV validation.
(DOC)

Acknowledgments

We thank the cell bank of Cochin and Salpétrière hospitals, the Plateforme de Ressource Biologique (B. Ghaleh), the Clinical Investigation Centers of Mondor-Chenevier, and Robert Debré hospitals for their assistance and the Psychiatry Genetics team (S. Jamain) of Inserm U955 for collection and management of DNAs. We thank the International Molecular Genetic Study of Autism Consortium (IMGSAC) for use of their samples. We thank C. Bouchier for the use of sequencing facilities and L. Lemée for the Illumina genotyping arrays at the Pasteur Génopole. We thank the Human Evolutionary Genetics laboratory of L. Quintana-Murci who provided us control DNA samples.

Author Contributions

Conceived and designed the experiments: T Bourgeron, CS Leblond, C Gillberg, TM Boeckers, J Heinrich, SW Scherer, D Pinto. Performed the experiments: CS Leblond, J Heinrich, D Pinto, G Huguet, B Oliveira, N Lemière, M Delepine, K Kantojärvi, A Chiocchetti. Analyzed the data: T Bourgeron, CS Leblond, TM Boeckers, J Heinrich, SW Scherer, D Pinto, G Huguet, E Ey. Contributed reagents/materials/analysis tools: C Proepper, G Huguet, M Konyukh, M Kopp, P Chaste, E Ey, M Rastam, H Anckarsäter, G Nygren, IC Gillberg, J Melke, R Toro, B Regnault, F Fauchereau, O Mercati, D Skuse, M Poot, R Holt, AP Monaco, I Järvelä, S Curran, DA Collier, P Bolton, SM Klauck, F Poustka, CM Freitag, E

Bacchelli, F Minopoli, A Battaglia, L Mazzone, E Maestrini, AF Sequeira, A Vicente, G Oliveira, D Zelenika, M Lathrop, V Guinchat, F Devillard, B Assouline, M-C Mouren, M Leboyer, C Gillberg, E Duketis, R Waltes, R Vanhala, D Bonneau, L Ruta. Wrote the paper: T Bourgeron, C Gillberg, M Leboyer, CS Leblond, R Delorme, C Betancur, R Toro.

References

1. American Psychiatric Association (1994) Diagnostic and Statistical Manual of Mental Disorders, 4th Ed American Psychiatric Press, Washington D.C.
2. Fernell E, Gillberg C (2010) Autism spectrum disorder diagnoses in Stockholm preschoolers. Res Dev Disabil 31: 680–685.
3. Freitag CM (2007) The genetics of autistic disorders and its clinical relevance: a review of the literature. Mol Psychiatry 12: 2–22.
4. Hallmayer J, Cleveland S, Torres A, Phillips J, Cohen B, et al. (2011) Genetic Heritability and Shared Environmental Factors Among Twin Pairs With Autism. Arch Gen Psychiatry.
5. Bourgeron T (2009) A synaptic trek to autism. Curr Opin Neurobiol 19: 231–234.
6. Toro R, Konyukh M, Delorme R, Leblond C, Chaste P, et al. (2010) Key role for gene dosage and synaptic homeostasis in autism spectrum disorders. Trends Genet: in press.
7. Durand CM, Betancur C, Boeckers TM, Bockmann J, Chaste P, et al. (2007) Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. Nat Genet 39: 25–27.
8. Jamain S, Quach H, Betancur C, Rastam M, Colineaux C, et al. (2003) Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. Nat Genet 34: 27–29.
9. Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, et al. (2010) Functional impact of global rare copy number variation in autism spectrum disorders. Nature 466: 368–372.
10. Szatmari P, Paterson AD, Zwaigenbaum L, Roberts W, Brian J, et al. (2007) Mapping autism risk loci using genetic linkage and chromosomal rearrangements. Nat Genet 39: 319–328.
11. Reddy KS (2005) Cytogenetic abnormalities and fragile-X syndrome in Autism Spectrum Disorder. BMC Med Genet 6: 3.
12. O'Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, et al. (2011) Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. Nat Genet 43: 585–589.
13. Sanders SJ, Ercan-Sencicek AG, Hus V, Luo R, Murtha MT, et al. (2011) Multiple Recurrent De Novo CNVs, Including Duplications of the 7q11.23 Williams Syndrome Region, Are Strongly Associated with Autism. Neuron 70: 863–885.
14. Levy D, Ronemus M, Yamrom B, Lee YH, Leotta A, et al. (2011) Rare de novo and transmitted copy-number variation in autistic spectrum disorders. Neuron 70: 886–897.
15. Poot M, Beyer V, Schwaab I, Damatova N, Van't Slot R, et al. (2010) Disruption of CNTNAP2 and additional structural genome changes in a boy with speech delay and autism spectrum disorder. Neurogenetics 11: 81–89.
16. Schaaf CP, Sabo A, Sakai Y, Crosby J, Muzny D, et al. (2011) Oligogenic heterozygosity in individuals with high-functioning autism spectrum disorders. Hum Mol Genet 20: 3366–3375.
17. Girirajan S, Rosenfeld JA, Cooper GM, Antonacci F, Siswara P, et al. (2010) A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay. Nat Genet 42: 203–209.
18. Betancur C (2011) Etiological heterogeneity in autism spectrum disorders: More than 100 genetic and genomic disorders and still counting. Brain Res 1380: 42–77.
19. Gilman SR, Iossifov I, Levy D, Ronemus M, Wigler M, et al. (2011) Rare de novo variants associated with autism implicate a large functional network of genes involved in formation and function of synapses. Neuron 70: 898–907.
20. Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, et al. (2011) A copy number variation morbidity map of developmental delay. Nat Genet 43: 838–846.
21. Sudhof TC (2008) Neuroligins and neurexins link synaptic function to cognitive disease. Nature 455: 903–911.
22. Yu LM, Goda Y (2009) Dendritic signalling and homeostatic adaptation. Curr Opin Neurobiol 19: 327–335.
23. Moessner R, Marshall CR, Sutcliffe JS, Skaug J, Pinto D, et al. (2007) Contribution of SHANK3 mutations to autism spectrum disorder. Am J Hum Genet 81: 1289–1297.
24. Gauthier J, Champagne N, Lafreniere RG, Xiong L, Spiegelman D, et al. (2010) De novo mutations in the gene encoding the synaptic scaffolding protein SHANK3 in patients ascertained for schizophrenia. Proc Natl Acad Sci U S A 107: 7863–7868.
25. Hamdan FF, Gauthier J, Araki Y, Lin DT, Yoshizawa Y, et al. (2011) Excess of de novo deleterious mutations in genes associated with glutamatergic systems in nonsyndromic intellectual disability. Am J Hum Genet 88: 306–316.
26. Berkel S, Marshall CR, Weiss B, Howe J, Roeth R, et al. (2010) Mutations in the SHANK2 synaptic scaffolding gene in autism spectrum disorder and mental retardation. Nat Genet 42: 489–491.
27. Persico AM, Bourgeron T (2006) Searching for ways out of the autism maze: genetic, epigenetic and environmental clues. Trends Neurosci 29: 349–358.
28. Boeckers TM, Kreutz MR, Winter C, Zuschratter W, Smalla KH, et al. (1999) Proline-rich synapse-associated protein-1/cortactin binding protein 1 (ProSAP1/CortBP1) is a PDZ-domain protein highly enriched in the postsynaptic density. J Neurosci 19: 6506–6518.
29. Sheng M, Kim E (2000) The Shank family of scaffold proteins. J Cell Sci 113(Pt 11): 1851–1856.
30. McWilliams RR, Gidey E, Fouassier L, Weed SA, Doctor RB (2004) Characterization of an ankyrin repeat-containing Shank2 isoform (Shank2E) in liver epithelial cells. Biochem J 380: 181–191.
31. Boeckers TM, Winter C, Smalla KH, Kreutz MR, Bockmann J, et al. (1999) Proline-rich synapse-associated proteins ProSAP1 and ProSAP2 interact with synaptic proteins of the SAPAP/GKAP family. Biochem Biophys Res Commun 264: 247–252.
32. Lim S, Naisbitt S, Yoon J, Hwang JI, Suh PG, et al. (1999) Characterization of the Shank family of synaptic proteins. Multiple genes, alternative splicing, and differential expression in brain and development. J Biol Chem 274: 29510–29518.
33. Durand CM, Kappeler C, Betancur C, Delorme R, Quach H, et al. (2006) Expression and genetic variability of PCDH11Y, a gene specific to Homo sapiens and candidate for susceptibility to psychiatric disorders. Am J Med Genet B Neuropsychiatr Genet 141: 67–70.
34. Gutierrez RC, Flynn R, Hung J, Kertesz AC, Sullivan A, et al. (2009) Activity-driven mobilization of post-synaptic proteins. Eur J Neurosci 30: 2042–2052.
35. Wischmeijer A, Magini P, Giorda R, Gnoli M, Ciccone R, et al. (2010) Olfactory Receptor-Related Duplicons Mediate a Microdeletion at 11q13.2q13.4 Associated with a Syndromic Phenotype. Mol Syndromol 1: 176–184.
36. Brandstatter JH, Dick O, Boeckers TM (2004) The postsynaptic scaffold proteins ProSAP1/Shank2 and Homer1 are associated with glutamate receptor complexes at rat retinal synapses. J Comp Neurol 475: 551–563.
37. Rosmond R, Bjorntorp P (1998) Psychiatric ill-health of women and its relationship to obesity and body fat distribution. Obes Res 6: 338–345.
38. Cloninger CR, Svrakic DM, Przybeck TR (1993) A psychobiological model of temperament and character. Arch Gen Psychiatry 50: 975–990.
39. Schalling D, Asberg M, Edman G, Oreland L (1987) Markers for vulnerability to psychopathology: temperament traits associated with platelet MAO activity. Acta Psychiatr Scand 76: 172–182.
40. Melke J, Westberg L, Nilsson S, Landen M, Soderstrom H, et al. (2003) A polymorphism in the serotonin receptor 3A (HTR3A) gene and its association with harm avoidance in women. Arch Gen Psychiatry 60: 1017–1023.
41. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, et al. (2010) A method and server for predicting damaging missense mutations. Nat Methods 7: 248–249.
42. Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, et al. (2008) Structural variation of chromosomes in autism spectrum disorder. Am J Hum Genet 82: 477–488.
43. Miller DT, Shen Y, Weiss LA, Korn J, Anselm I, et al. (2009) Microdeletion/duplication at 15q13.2q13.3 among individuals with features of autism and other neuropsychiatric disorders. J Med Genet 46: 242–248.
44. Depienne C, Moreno-De-Luca D, Heron D, Bouteiller D, Gennetier A, et al. (2009) Screening for genomic rearrangements and methylation abnormalities of the 15q11–q13 region in autism spectrum disorders. Biol Psychiatry 66: 349–359.
45. van der Zwaag B, Staal WG, Hochstenbach R, Poot M, Spierenburg HA, et al. (2010) A co-segregating microduplication of chromosome 15q11.2 pinpoints two risk genes for autism spectrum disorder. Am J Med Genet B Neuropsychiatr Genet 153B: 960–966.
46. Shen Y, Dies KA, Holm IA, Bridgemohan C, Sobeih MM, et al. (2010) Clinical genetic testing for patients with autism spectrum disorders. Pediatrics 125: e727–735.
47. Sharp AJ, Mefford HC, Li K, Baker C, Skinner C, et al. (2008) A recurrent 15q13.3 microdeletion syndrome associated with mental retardation and seizures. Nat Genet 40: 322–328.
48. Masurel-Paulet A, Andrieux J, Callier P, Cuisset JM, Le Caignec C, et al. (2010) Delineation of 15q13.3 microdeletions. Clin Genet 78: 149–161.
49. Doornbos M, Sikkema-Raddatz B, Ruijvenkamp CA, Dijkhuizen T, Bijlsma EK, et al. (2009) Nine patients with a microdeletion 15q11.2 between breakpoints 1 and 2 of the Prader-Willi critical region, possibly associated with behavioural disturbances. Eur J Med Genet 52: 108–115.
50. Helbig I, Mefford HC, Sharp AJ, Guipponi M, Fichera M, et al. (2009) 15q13.3 microdeletions increase risk of idiopathic generalized epilepsy. Nat Genet 41: 160–162.
51. Szafranski P, Schaaf CP, Person RE, Gibson IB, Xia Z, et al. (2010) Structures and molecular mechanisms for common 15q13.3 microduplications involving CHRNA7: benign or pathological? Hum Mutat 31: 840–850.
52. Ben-Shachar S, Lanpher B, German JR, Qasaymeh M, Potocki L, et al. (2009) Microdeletion 15q13.3: a locus with incomplete penetrance for autism, mental retardation, and psychiatric disorders. J Med Genet 46: 382–388.

53. Stefansson H, Rujescu D, Cichon S, Pietilainen OP, Ingason A, et al. (2008) Large recurrent microdeletions associated with schizophrenia. Nature 455: 232–236.
54. International Schizophrenia Consortium (2008) Rare chromosomal deletions and duplications increase risk of schizophrenia. Nature 455: 237–241.
55. Tam GW, van de Lagemaat LN, Redon R, Strathdee KE, Croning MD, et al. (2010) Confirmed rare copy number variants implicate novel genes in schizophrenia. Biochem Soc Trans 38: 445–451.
56. Murthy SK, Nygren AO, El Shakankiry HM, Schouten JP, Al Khayat AI, et al. (2007) Detection of a novel familial deletion of four genes between BP1 and BP2 of the Prader-Willi/Angelman syndrome critical region by oligo-array CGH in a child with neurological disorder and speech impairment. Cytogenet Genome Res 116: 135–140.
57. Shinawi M, Schaaf CP, Bhatt SS, Xia Z, Patel A, et al. (2009) A small recurrent deletion within 15q13.3 is associated with a range of neurodevelopmental phenotypes. Nat Genet 41: 1269–1271.
58. van Bon BW, Mefford HC, Menten B, Koolen DA, Sharp AJ, et al. (2009) Further delineation of the 15q13 microdeletion and duplication syndromes: a clinical spectrum varying from non-pathogenic to a severe outcome. J Med Genet 46: 511–523.
59. Pagnamenta AT, Wing K, Akha ES, Knight SJ, Bolte S, et al. (2009) A 15q13.3 microdeletion segregating with autism. Eur J Hum Genet 17: 687–692.
60. Kirov G, Grozeva D, Norton N, Ivanov D, Mantripragada KK, et al. (2009) Support for the involvement of large copy number variants in the pathogenesis of schizophrenia. Hum Mol Genet 18: 1497–1503.
61. de Kovel CG, Trucks H, Helbig I, Mefford HC, Baker C, et al. (2010) Recurrent microdeletions at 15q11.2 and 16p13.11 predispose to idiopathic generalized epilepsies. Brain 133: 23–32.
62. Cottrell CE, Bir N, Varga E, Alvarez CE, Bouyain S, et al. (2010) Contactin 4 as an autism susceptibility locus. Autism Res.
63. Rosenfeld JA, Ballif BC, Torchia BS, Sahoo T, Ravnan JB, et al. (2010) Copy number variations associated with autism spectrum disorders contribute to a spectrum of neurodevelopmental disorders. Genet Med 12: 694–702.
64. Christian SL, Brune CW, Sudi J, Kumar RA, Liu S, et al. (2008) Novel submicroscopic chromosomal abnormalities detected in autism spectrum disorder. Biol Psychiatry 63: 1111–1117.
65. Vrijenhoek T, Buizer-Voskamp JE, van der Stelt I, Strengman E, Sabatti C, et al. (2008) Recurrent CNVs disrupt three candidate genes in schizophrenia patients. Am J Hum Genet 83: 504–510.
66. Yamakawa H, Oyama S, Mitsuhashi H, Sasagawa N, Uchino S, et al. (2007) Neuroligins 3 and 4X interact with syntrophin-gamma2, and the interactions are affected by autism-related mutations. Biochem Biophys Res Commun 355: 41–46.
67. Alessi A, Bragg AD, Percival JM, Yoo J, Albrecht DE, et al. (2006) gamma-Syntrophin scaffolding is spatially and functionally distinct from that of the alpha/beta syntrophins. Exp Cell Res 312: 3084–3095.
68. Vierbuchen T, Ostermeier A, Pang ZP, Kokubu Y, Sudhof TC, et al. (2010) Direct conversion of fibroblasts to functional neurons by defined factors. Nature 463: 1035–1041.
69. Durand CM, Perroy J, Loll F, Perrais D, Fagni L, et al. (2011) SHANK3 mutations identified in autism lead to modification of dendritic spine morphology via an actin-dependent mechanism. Mol Psychiatry.
70. Berkel S, Tang W, Trevino M, Vogt M, Obenhaus HA, et al. (2011) Inherited and de novo SHANK2 variants associated with autism spectrum disorder impair neuronal morphogenesis and physiology. Hum Mol Genet.
71. Durbin RM, Abecasis GR, Altshuler DL, Auton A, Brooks LD, et al. (2010) A map of human genome variation from population-scale sequencing. Nature 467: 1061–1073.
72. The international schizophrenia consortium (2008) Rare chromosomal deletions and duplications increase risk of schizophrenia. Nature 455: 237–241.
73. Yasui DH, Scoles HA, Horike SI, Meguro-Horike M, Dunaway KW, et al. (2011) 15q11.2–13.3 chromatin analysis reveals epigenetic regulation of CHRNA7 with deficiencies in Rett and autism brain. Hum Mol Genet.
74. Meguro-Horike M, Yasui DH, Powell W, Schroeder DI, Oshimura M, et al. (2011) Neuron-specific impairment of inter-chromosomal pairing and transcription in a novel model of human 15q-duplication syndrome. Hum Mol Genet 20: 3798–3810.
75. Hogart A, Leung KN, Wang NJ, Wu DJ, Driscoll J, et al. (2009) Chromosome 15q11–13 duplication syndrome brain reveals epigenetic alterations in gene expression not predicted from copy number. J Med Genet 46: 86–93.
76. Triana-Baltzer GB, Liu Z, Gounko NV, Berg DK (2008) Multiple cell adhesion molecules shaping a complex nicotinic synapse on neurons. Mol Cell Neurosci 39: 74–82.
77. Cheng SB, Amici SA, Ren XQ, McKay SB, Treuil MW, et al. (2009) Presynaptic targeting of alpha4beta 2 nicotinic acetylcholine receptors is regulated by neurexin-1beta. J Biol Chem 284: 23251–23259.
78. Neff RA, 3rd, Gomez-Varela D, Fernandes CC, Berg DK (2009) Postsynaptic scaffolds for nicotinic receptors on neurons. Acta Pharmacol Sin 30: 694–701.
79. Schenck A, Bardoni B, Langmann C, Harden N, Mandel JL, et al. (2003) CYFIP/Sra-1 controls neuronal connectivity in Drosophila and links the Rac1 GTPase pathway to the fragile X protein. Neuron 38: 887–898.
80. Kelleher RJ, 3rd, Bear MF (2008) The autistic neuron: troubled translation? Cell 135: 401–406.
81. Boeckers TM, Bockmann J, Kreutz MR, Gundelfinger ED (2002) ProSAP/Shank proteins - a family of higher order organizing molecules of the postsynaptic density with an emerging role in human neurological disease. J Neurochem 81: 903–910.
82. Sugiyama Y, Kawabata I, Sobue K, Okabe S (2005) Determination of absolute protein numbers in single synapses by a GFP-based calibration technique. Nat Methods 2: 677–684.
83. Nagai R, Hashimoto R, Tanaka Y, Taguchi O, Sato M, et al. (2010) Syntrophin-2 is required for eye development in Drosophila. Exp Cell Res 316: 272–285.
84. Dolen G, Osterweil E, Rao BS, Smith GB, Auerbach BD, et al. (2007) Correction of fragile X syndrome in mice. Neuron 56: 955–962.
85. Vissers LE, de Ligt J, Gilissen C, Janssen I, Steehouwer M, et al. (2010) A de novo paradigm for mental retardation. Nat Genet 42: 1109–1112.
86. Colella S, Yau C, Taylor JM, Mirza G, Butler H, et al. (2007) QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. Nucleic Acids Res 35: 2013–2025.
87. Wang K, Li M, Hadley D, Liu R, Glessner J, et al. (2007) PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res 17: 1665–1674.
88. Boeckers TM, Liedtke T, Spilker C, Dresbach T, Bockmann J, et al. (2005) C-terminal synaptic targeting elements for postsynaptic density proteins ProSAP1/Shank2 and ProSAP2/Shank3. J Neurochem 92: 519–524.

J. Pineal Res. 2013; 54:46–57
Doi:10.1111/j.1600-079X.2012.01020.x

Journal of Pineal Research

Journal of Pineal Research
Molecular, Biological, Physiological and Clinical Aspects of Melatonin

Crystal structure and functional mapping of human ASMT, the last enzyme of the melatonin synthesis pathway

Abstract: Melatonin is a synchronizer of many physiological processes. Abnormal melatonin signaling is associated with human disorders related to sleep, metabolism, and neurodevelopment. Here, we present the X-ray crystal structure of human *N*-acetyl serotonin methyltransferase (ASMT), the last enzyme of the melatonin biosynthesis pathway. The polypeptide chain of ASMT consists of a C-terminal domain, which is typical of other SAM-dependent *O*-methyltransferases, and an N-terminal domain, which intertwines several helices with another monomer to form the physiologically active dimer. Using radioenzymology, we analyzed 20 nonsynonymous variants identified through the 1000 genomes project and in patients with neuropsychiatric disorders. We found that the majority of these mutations reduced or abolished ASMT activity including one relatively frequent polymorphism in the Han Chinese population (N17K, rs17149149). Overall, we estimate that the allelic frequency of ASMT deleterious mutations ranges from 0.66% in Europe to 2.97% in Asia. Mapping of the variants on to the 3-dimensional structure clarifies why some are harmful and provides a structural basis for understanding melatonin deficiency in humans.

Hany Goubran Botros[1,2,3*], Pierre Legrand[4*], Cecile Pagan[1,2,3], Vincent Bondet[5], Patrick Weber[6], Mariem Ben-Abdallah[5], Nathalie Lemière[1,2,3], Guillaume Huguet[1,2,3], Jacques Bellalou[5], Erik Maronde[7], Pierre Beguin[5], Ahmed Haouz[6], William Shepard[4,6] and Thomas Bourgeron[1,2,3]

[1]Institut Pasteur, Human Genetics and Cognitive Functions Unit, Paris, France; [2]CNRS URA 2182 'Genes, synapses and cognition', Institut Pasteur, Paris, France; [3]University Paris Diderot, Sorbonne Paris Cité, Human Genetics and Cognitive Functions, Paris, France; [4]Synchrotron SOLEIL, L'Orme des Merisiers, Saint Aubin BP48, Gif-sur-Yvette, France; [5]Institut Pasteur, Plate forme 5, 25 rue Dr. Roux, Paris, France; [6]Institut Pasteur, Plate forme 6, CNRS-UMR3528, 25 rue Dr. Roux, Paris, France; [7]Institute for Anatomy III, Goethe University, Frankfurt, Germany

Key words: ASMT, Genetic variants, crystal structure, melatonin, O-methyl transferase

Address reprint requests to Thomas Bourgeron, PhD, Human Genetics and Cognitive Functions, Institut Pasteur, 25 rue du Docteur Roux, 75015 Paris, France.
E-mail: thomasb@pasteur.fr or
William Shepard, PhD, Synchrotron SOLEIL, L'Orme des Merisiers, Saint Aubin BP48, 91192 Gif-sur-Yvette, France.
E-mail: william.shepard@synchrotron-soleil.fr.
*These authors contributed equally to this work.

Received April 21, 2012;
Accepted June 1, 2012.

Introduction

Melatonin is a multitasking molecule synthesized in the pineal gland during the night and involved in various physiological functions including sleep induction, circadian rhythm regulation, as well as oxidative stress and immune response [1]. In humans, melatonin levels are heterogeneous among individuals and under strong genetic influence [2, 3]. In some individuals, the endogenous level of this hormone might not be sufficient to set their biological clock, and abnormal melatonin signaling has been reported as a risk factor for diverse medical conditions, such as diabetes and psychiatric disorders [4]. In mice, melatonin suppresses *Clock* mutant phenotypes and affects the circadian system [5]. Melatonin synthesis requires serotonin, which is first acetylated by the arylalkylamine *N*-acetyltransferase (AA-NAT) to produce *N*-acetyl serotonin (NAS) (Fig. 1A). Then, acetyl serotonin methyltransferase (ASMT, also known as hydroxyindole O-methyltransferase or HIOMT) produces melatonin by transferring a methyl group from the cofactor *S*-adenosyl-L-methionine (SAM) to NAS [6]. Variants of ASMT were identified in the human population, and some are associated with dramatic decreases in melatonin synthesis in patients with neuropsychiatric

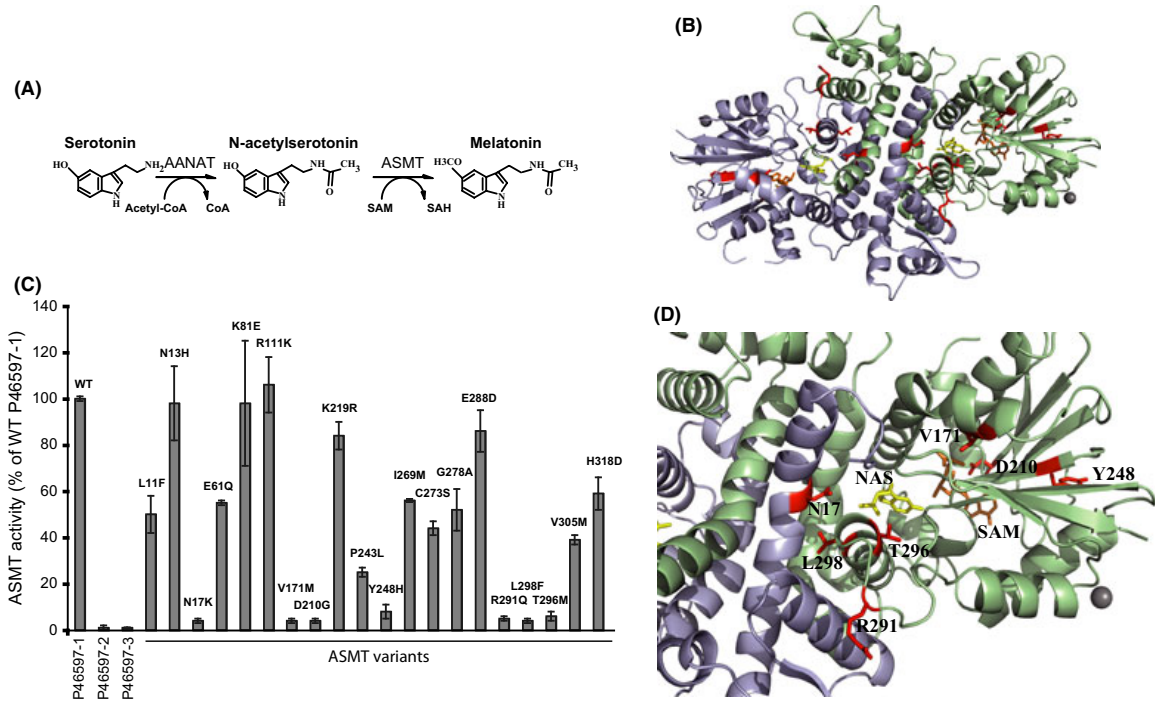


Fig. 1. The crystal structure of the human acetyl serotonin methyltransferase (ASMT) and the impact of genetic variants on ASMT activities. (A) Melatonin biosynthetic pathway. The enzymes depicted are AANAT, arylalkylamine *N*-acetyltransferase, and ASMT, acetyl serotonin methyltransferase. (B) Ribbon diagram of the crystal structure of ASMT bound to NAS and *S*-adenosyl-L-methionine (SAM) (PDB codes 4A6D and 4A6E). (C) Functional impact of ASMT variants on the enzyme activity. Among the three ASMT isoforms (P46597-1, P46597-2, and P46597-3), only P46597-1 is able to synthesize melatonin. Numbering of ASMT variants is based on the sequence P46597-1, the protein sequence coding for the functional isoform of ASMT. The wild-type ASMT enzyme activity was 21 ± 2 nmol/h/mg protein. (D) Detailed view of the active site of ASMT interacting with *N*-acetyl serotonin (NAS) and SAM. Human variants with a strong impact on ASMT activity are indicated in red.

disorders such as autism spectrum disorders (ASD), attention-deficit hyperactivity disorders (ADHD), or intellectual disability (ID) [7–11]. Recently, we have cloned and expressed human ASMT as a soluble recombinant protein in *Leishmania tarentolae* [12], and we confirmed the presence of the protein in the human pineal gland by immunofluorescence [13].

The biosynthetic pathway of melatonin has been known for 50 years, and the crystal structure describing each enzyme involved in this pathway starting from L-tryptophan has been identified (Protein Data Bank (PDB) ID codes: 1MLW, 3RBF, 1KUX), except for the last step catalyzed by ASMT. To better understand the impact of ASMT mutations on melatonin synthesis, we undertook the resolution of the crystal structure of human ASMT, and we mapped the binding sites for both NAS and SAM. By comparing the functional effect of a large panel of 20 genetic variants with the 3-dimensional structure, these results provide the first structural basis of ASMT deficiency in humans.

Material and methods

Crystallization strategy

Initial identification of crystallization conditions was carried out using the vapor diffusion method in a Cartesian technology workstation. Sitting drops were set using 400 nL of 1:1 mixture of ASMT protein and a crystallization solution (672 different conditions, commercially available), equilibrating against 150-μL reservoir in a Greiner plate [14]. Optimization of initial hits was pursued manually in Linbro plates with a hanging drop setup. The best crystals were obtained by mixing 1.5 μL of native ASMT protein at 14 mg/mL with 1.5 μL of the reservoir solution containing 30% (V/V) PEG-400, 0.1 M cacodylate pH 6.5, and 0.2 M lithium sulfate at 18°C. The crystals appeared within one week and had dimensions of up to 0.1 mm × 0.1 mm × 0.2 mm. Single crystals of the ASMT protein were flash-frozen in liquid nitrogen using the crystallization solution as cryoprotectant.

Collection and processing of X-ray diffraction data

X-ray diffraction data were collected from native and heavy-atom derivative crystals of ASMT on the beamlines PROXIMA 1 at SOLEIL and ID23-EH2 at the ESRF (Table 1). Initial attempts to solve the crystal structure via molecular replacement methods failed because a dimer was assumed to occupy the asymmetric unit. A crystal soaked in para-chloromercury-benzoate (PCMB) diffracting to 3.3 Å resolution was used to solve the structure by the single-wavelength anomalous dispersion (SAD) method. The diffraction images

Table 1. Crystallographic parameters and data statistics

Crystal name	acetyl serotonin methyltransferase (ASMT) with (PCMB)	ASMT + (*S*-adenosyl-L-methionine (SAM) + *N*-acetyl serotonin)	ASMT + (SAM)
X-ray source	SOLEIL-Proxima1	ESRF-ID23-2	SOLEIL-Proxima1
X-ray wavelength (Å)	1.0060	0.8726	1.0720
Space group	*R32*	*R32*	*R32*
Unit-cell dimensions (Å)	$a = b = 173.6$ $c = 107.3$	$a = b = 170.1$ $c = 128.2$	$a = b = 170.4$ $c = 123.3$
Data resolution (Å)	43.7–3.5 (3.7–3.5)	42.7–2.7 (2.86–2.7)	49.2–2.4 (2.55–2.4)
Unique reflections	15234 (2372)	19684 (3141)	26703 (4104)
Completeness (%)	99.2 (95.3)	99.7 (98.0)	99.1 (95.0)
Redundancy	5.6 (5.5)	7.5 (7.5)	16.1 (10.8)
Rmerge	0.084 (0.424)	0.086 (0.750)	0.062 (0.857)
$< I/\sigma(I) >$	16.2 (4.2)	23.3 (3.4)	26.6 (2.8)
Anomalous correlation	0.42 (0.23 *)	–	–
Anomalous ratio (rms($\Delta I \pm$)/rms(I))	1.30 (1.06 *)	–	–
Phasing statistics (PHASER)			
Resolution limits	43.6–3.5	–	–
FOM (acentrics/centrics)	0.320/0.152	–	–
Refinement statistics			
Protein Data Bank ID code	–	4A6E	4A6D
Resolution (Å)	–	42.7–2.70 (2.85–2.7)	41.4–2.4 (2.5–2.4)
R-factor (working set)	–	0.1633 (0.2192)	0.1759 (0.2353)
Rfree	–	0.2122 (0.2513)	0.2041 (0.2803)
No. of unique reflections (total)	–	19662 (2828)	26695 (2827)
No. of protein atoms	–	2697	2712
No. of water molecules	–	49	67
No. of ions and ligand atoms	–	129	122
Geometry: rms deviations from ideal values			
Bond lengths (Å)	–	0.010	0.010
Bond angles (°)	–	1.13	1.12
Mean protein B factor values (Å^2)	–	69.9	86.0

The unit-cell dimensions are given in their hexagonal equivalents for the space group R32. Values in parentheses apply to the high-resolution shells as indicated, except for those marked with asterisks (*), which are 4.7–4.3 Å for the anomalous correlation and anomalous ratio. All statistical parameters are as defined in the programs XDS, PHASER, and BUSTER. PCMB, para-chloro-mercury-benzoate.

were processed and scaled with XDS [15]. The mercury sites were found with SHELXD [16], and the initial map was calculated with PHASER [17] and solvent-flattened with PARROT [18]. Buccaneer [19] traced a single protein molecule in the asymmetric unit. The coordinates of this structure were employed to determine a molecular replacement solution with MOLREP [20] of the SAM bound form (2.4 Å resolution) and the SAM + NAS bound form (2.7 Å resolution). The structure was refined with BUSTER-TNT [21] and rebuilt with COOT [22]. Calculation of the homodimer interface area was completed using PISA [23]. An unknown metal ion was discovered at the surface of the protein and identified as a zinc ion by collecting diffraction data at energies above and below the Zn K-edge (Fig. 1B,D). The presence of this zinc ion is most likely an artifact from residual zinc bound to the His-tag of the recombinant protein. Coordinates for the native ASMT forms with SAM and SAM + NAS have been deposited in the Protein Data Bank: PDB ID codes 4A6D and 4A6E, respectively.

Construction of mutagenized cDNAs encoding human ASMT

Human cDNA encoding *ASMT* was cloned in a PcDNA-dest47 vector. Site-directed mutagenesis was performed with the QuickChange II XL Site Directed Mutagenesis Kit (Invitrogen Life Technologies, Grand Island, NY, USA). Each clone was purified with Endofree Plasmid Maxi Kit (Qiagen, Hilden, Germany) and entirely sequenced to rule out additional mutations in the cDNA. COS-7 cells were transiently transfected using the Amaxa Nucleofector Kit (Lonza) according to the manufacturer's instructions. cDNAs were cotransfected with luciferase as a reporter gene for ascertaining transfection efficiency. We constructed *ASMT* mutants with the following substitutions that we previously identified: L11F, N13H, N17K, E61Q, K81E, R111K, V171M, D210G, K219R, Y248H, I269M, P243L, C273S, G278A, E288D, R219Q, V305M, T296M, L298F, and H318D. The sequence of all constructs was confirmed by direct sequencing.

Cell and tissue homogenate preparation

Cell homogenates of cultured ASMT-COS-7 cells were prepared by scraping in the presence of 4% of the zwitterionic detergent CHAPS and pelleted by centrifugation (15 min, 13,000 g, at 4°C). Human postmortem pineal glands were sonicated on ice in sodium phosphate buffer (0.05 M, pH 7.9), and the sonicate was centrifuged at 15,000 g for 20 min. The resulting supernatant fractions of both tissue and cell homogenate were subjected to Western blotting and chromatography and assayed for ASMT

activity. A small amount of the supernatant was used for determination of protein concentration by BCA protein test (ThermoScientific, Waltham, MA, USA).

Recombinant human ASMT was produced and purified using *Leishmania tarentolae* as previously described [12]. The apparent molecular weight of the monomer of the purified enzyme as well as the enzyme from homogenates of ASMT-transfected COS-7 cells and human pineal glands was ~39 kDa, as determined by SDS-PAGE (Fig. 2). However, in all preparations, ASMT activity was detected in the fractions corresponding to the dimeric form.

Size-exclusion chromatography and radiometric assay of ASMT activity

Size-exclusion chromatography was carried out on Superdex 75 10–300 GL column (GE Healthcare) equilibrated with 0.05 M phosphate buffer, pH 7.0, containing 0.05 M NaCl. ASMT activity was measured as described [24, 25]. A 50-μL sample of homogenate (ASMT-transfected COS-7 or pineal glands) in 0.05 M sodium phosphate buffer, pH 7.9, was incubated (37°C; 30 min) with 25 μL of *N*-acetylserotonin (NAS) (final concentration, 50 μM) and 25 μL [^{3}H-methyl]-*S*-adenosyl-L-methionine (SAM) (final concentration 300 μM; specific activity 50 Ci/mol) in a total volume of 100 μL. Enzymatic reactions were stopped by adding 200 μL 0.45 M sodium borate buffer, pH 10, and 1 mL chloroform. The radiolabelled [^{3}H] melatonin containing organic phase was washed twice with borate buffer. Finally, 0.8 mL samples of the organic phase were evaporated to dryness, and radioactivity was determined using a scintillation counter. Background was calculated as the radioactive products formed in the absence of NAS. The impact of the ASMT variants on function was predicted *in silico* using six different algorithms: Polyphen2 (http://genetics.bwh.harvard.edu/pph2/), nsSNPAnalyzer (http://snpanalyzer.uthsc.edu/), PMUT (http://mmb.pcb.ub.es/PMut/PMut.jsp), SNAP (http://rostlab.org/services/snap/submit), SIFT (http://sift.jcvi.org/), and PANTHER (http://www.pantherdb.org/tools/csnpScoreForm.jsp).

The rabbit anti-human ASMT antibody and Western blot analysis

The antiserum raised in rabbits against a mixture of two human synthetic ASMT peptides (H2N-GSSED-QAYRLLNDYAC-CONH2 and H2N-CQHFSFQEEE-QIDFQE-CONH2) was previously used to characterize the morphology of the human pineal gland [13]. Here, we used clarified lysates of the transfected COS cells (3 μg) or of the human pineal glands (1 μg). Proteins were resolved on a 12% precast gels containing lithium dodecyl sulfate (LDS) under reducing conditions (NuPAGE; Invitrogen) and electrotransferred onto polyvinylidene difluoride membrane (iBlot system, Life Technologies) according to the manufacturer's protocol. Immunodetection was carried out using rabbit anti-ASMT antiserum (Eurogentec), and a mouse monoclonal anti-ß-actin antibody (Sigma-Aldrich, St. Louis, MO, USA) was used for normalization. The semi-quantification densitometric analysis of the ASMT-positive bands in the Western blots was carried out using the Odyssey Infrared Imaging System (LI-COR, Lincoln, NE, USA). After transfection, cell homogenates were subjected to Western blotting detection and the levels of ASMT were comparable for all variants.

Results

The recombinant human ASMT purified from *Leishmania tarentolae* [12] was crystallized in the space group R32 (space group symmetries of crystals are described in the International Tables for Crystallography Volume A) with a single protein molecule in the asymmetric unit (Table 1). The native crystals, bound with SAM alone and bound with both SAM and NAS, diffracted to 2.4 and 2.7 Å resolutions, respectively. The first crystal structure was determined using the SAD method from a single mercury atom derivative. The resulting phased map revealed clear electron density for the entire protein molecule except for two residues (225–226). The polypeptide chain of human ASMT comprises two domains. The C-terminal domain (residues

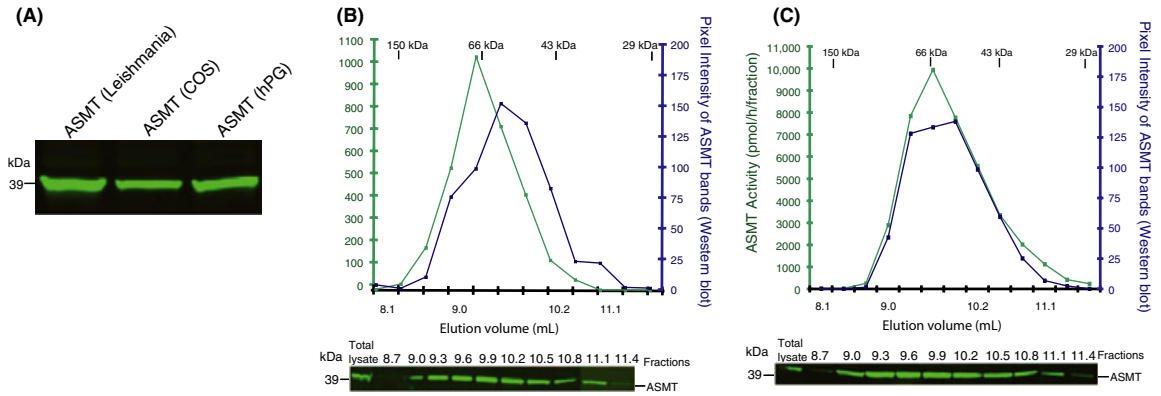


Fig. 2. Biochemical characterization of the human ASMT protein. (A) Western blot of the recombinant human ASMT (rhASMT) isolated from *Leishmania tarentolae*, rhASMT isolated from transfected COS cells, and the endogenous ASMT in homogenates from human pineal gland (hPG) using anti-human ASMT antibody. (B, C) Size-exclusion chromatography analysis of rhASMT from transfected COS cells (B) and ASMT from hPG (C). Collected fractions were analyzed for their content of ASMT using anti-ASMT antibodies (blue lane) as well as for ASMT activity (green lane). ASMT activity was bound to fractions corresponding to molecular mass of dimeric form.

141–288 and 307–347) is typical of other SAM-dependent *O*-methyltransferases (O-MTs) [26], that is, a set of alternating beta strands and alpha helices, while the N-terminal domain (residues 1–140 and 289–306) intertwines several helices into a bundle with another crystallographically equivalent monomer to form the physiologically active dimer (Fig. 1B).

The overall fold of the ASMT dimer closely resembles the plant O-MTs: chalcone O-MT and isoflavone O-MT [26, 27]. A search of the PDB archives with PDBeFold [28] reveals over 50 structures from at least 15 proteins with folds similar to the ASMT structures (rms deviations of 2.5–3.8Å). Surprisingly, these proteins, which come only from plants and bacteria, have sequence identities of 26% or less with human ASMT. It is interesting to note that when compared to this subfamily of SAM-dependant O-MTs, the ASMT structures presented here are in a 'closed' conformation, such that the C-terminal and N-terminal domains bring the SAM cofactor and the NAS substrate closely together. This situation is akin to the structure of isoflavone O-MT from alfalfa (PDB ID: 1fp2), but very different from the 'open' configurations seen for the chalcone O-MT from alfalfa (PDB ID: 1fp1), isoflavone O-MT from barrel medic (PDB ID: 1zga and 2qyo), caffeic acid O-MT from alfalfa (PDB ID: 1kyz), and others [29, 30].

The C-terminal domain of ASMT is very similar to those of small-molecule MTs, which are monomeric proteins, such as catechol O-MT [31]. Some of these small-molecule MTs contain a divalent metal ion in the active site, but this is not the case for ASMT. The methyl donor SAM resides in a well-buried pocket formed by four loops and a helix. It makes the expected H-bonds to D210 and Y147 from the hydroxyl groups of the sugar moiety, to G187 and R252 from the amino acid end, and to G235 and D236 from the adenosyl group, with the latter being packed between I211 and F237 (Fig. 3). When soaked with 20 mM of NAS and 20 mM of SAM, the crystal structure reveals the substrate NAS molecule bound into a pocket adjacent to the SAM molecule. The NAS molecule forms H-bonds to the residues H255, D256, N302, Q306, and two water molecules (Fig. 3). It is also packed by several hydrophobic residues including F156, L160, Y108, Y299, and M303. One side of the NAS site borders the dimer interface near residues N17 and M20. Interestingly, the thermal parameters of the residues surrounding the NAS pocket are higher in the absence of NAS, especially the loop F135-E150, suggesting that this zone is flexible to permit the entry and exit of the substrate (Fig. 4).

Bridging the SAM and NAS are the residues H255, D256, and the main-chain carbonyl group of R252. The bound SAM and NAS molecules are orientated by their respective binding sites so that the methyl group of the SAM moiety and the hydroxyl group of the indole group of NAS are aligned and brought close together. In fact, both substrates are perfectly poised for methyl transfer (Fig. 3). The methyl group transfer is presumably activated by abstraction of a proton from the hydroxyl of NAS by the imidazole group of H255 assisted by the carboxylate group of E311, a mechanism similar to those proposed for other O-MTs [32].

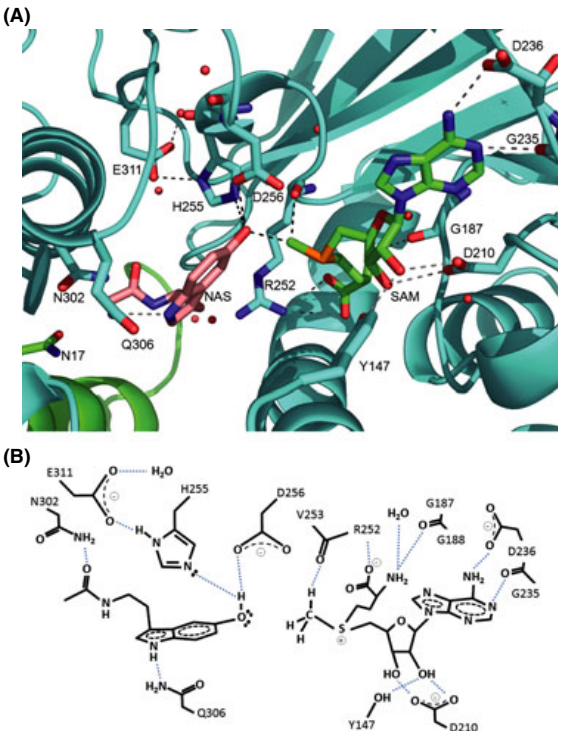


Fig. 3. Key residues of the active site of acetyl serotonin methyltransferase (ASMT). (A) Close-up of the active site of ASMT (cyan) with *N*-acetyl serotonin (NAS) (pink) and *S*-adenosyl-L-methionine (SAM) (green). The carbonyl of R252 and the side chains of H255 and D256 are situated just above the activated methyl group. Certain parts of the active site resemble those of many other O-MTs. (B) Schematic diagram of the active site of ASMT indicating residues that make H-bonds to the SAM and NAS molecules.

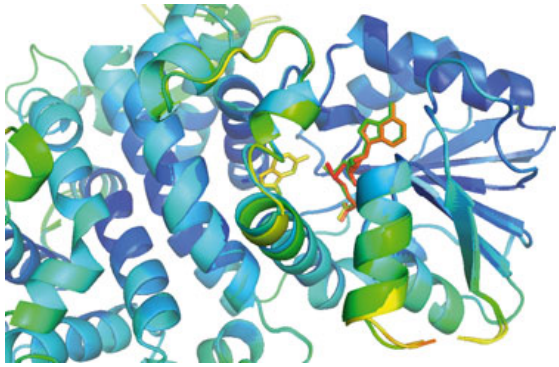


Fig. 4. Changes in thermal parameters of human acetyl serotonin methyltransferase (ASMT) upon binding *N*-acetyl serotonin (NAS) suggest entry portals to binding sites. Ribbon diagram of ASMT bound to *S*-adenosyl-L-methionine (SAM) and SAM + NAS colored according to thermal parameters of the crystal structures: low thermal parameters in blue and high thermal parameters in yellow/orange.

The dimerization domain of ASMT has at its core a bundle of four (2 × 2) long alpha helices, which are bent and tightly intertwined. This bundle is surrounded by several other shorter helices plus two short beta strands located at the extremities. This domain also plugs the *N*-acetyl side of the NAS binding site of an opposing monomer, where the N17 side-chain forms a water-mediated H-bond with the NAS substrate. In fact, this residue is located in the bend of a long helix (α1), which makes intimate contacts along the dimerization surface and effectively blocks the access of the substrate site to solvent. The dimerization interface is very extensive and covers 3545 Å^2. It is mostly comprised of hydrophobic interactions and a few polar contacts, all of which stabilize the homodimer.

When compared to the amino acid sequences of 13 other proteins identified as ASMT from other organisms, many residues in ASMT are well, if not strictly, conserved

```
SP|P46597|ASMT_HUMAN    ------MG-----SSEDQAY-RLLNDYANGFMVSQVLFAACELGVFDLLAEAPGPLDVAA 48
SP|Q8HZJ0|ASMT_MACMU    ------MG-----SSGDDGY-RLLNEYTNGFMVSQVLFAACELGVFDLLAEAPGPLDVAA 48
SP|P10950|ASMT_BOVIN    ------MC-----SQEGEGY-SLLKEYANGFMVSQVLFAACELGVFELLAEALEPLDSAA 48
TR|F6WEJ3|F6WEJ3_HORSE  ------MS-----SLEDQAF-GLLNTYSNGFMVSQVLFAACELGVFDLLAEAREPLDSAV 48
TR|F6X3D5|F6X3D5_XENTR LAVQLNMS-----SVQDLDYPKQLLEYRDGFLFSKTLFTACELGIFDLLHKSGEALSTLA 55
TR|G3NTU3|G3NTU3_GASAC MAESGEMMGPSEGTPAADTFPKKILEYMEGFLVSKTVFSACELGVFDVLLAADHPLSAEE 60
SP|Q92056|ASMT_CHICK    ------MD-----STEDLDYPQIIFQYSNGFLVSKVMFTACELGVFDLLLQSGRPLSLDV 49
TR|G1NPG2|G1NPG2_MELGA  ----RKMD-----STEDLDYPQIIFQYSNGFLVSKVMFTACELGVFDLLLESGQPLSLDV 51
TR|G3WBA1|G3WBA1_SARHA  ------MS-----FAEDLDY-DLFNEYINGFMISKVMFTACELGVFDLLESEG-FLTFDV 47
SP|D3KU67|ASMT_MUSMM    ------MHRGRSASARQERDFRALMDLAHGFMASQVLFAGCALRVFDAA--ALGPVDAAA 52
SP|D3KU66|ASMT_MOUSE    ------MHRGRSASARQERDFRALMDLAHGFMASQVLFAGCALRVFDAA--ALGPVDAAA 52
SP|B3GSH5|ASMT_RAT      ------MAPGR--EGELDRDFRVLMSLAHGFMVSQVLFAALDLGIFDLA--AQGPVAAEA 50
TR|E9PSX5|E9PSX5_RAT    ------MAPGR--EGELDRDFRVLMSLAHGFMVSQVLFAALDLGIFDLA--AQGPVAAEA 50
                              *                :    .**: *:.:*:.  * :*:        :

SP|P46597|ASMT_HUMAN    VAAGVRASAHGTELLLDICVSLKLLKVETRGG--KAFYRNTELSSDYLTTVSPTSQCSML 106
SP|Q8HZJ0|ASMT_MACMU    VAAGVEASSHGTELLLDTCVSLKLLKVETRAG--KAFYQNTELSSAYLTRVSPTSQCNLL 106
SP|P10950|ASMT_BOVIN    VSSHLGSSPQGTELLLNTCVSLKLLQADVRGG--KAVYANTELASTYLVRGSPRSQRDML 106
TR|F6WEJ3|F6WEJ3_HORSE VAARLGASCHGTELLLDACVTLKLLRVDTRRG--KAFYGNTELSSTYLTTASPTTQRNML 106
TR|F6X3D5|F6X3D5_XENTR ISSQLGTNTDATDRLLSACVGLKLLKVEMKNN--EAFFSNTDVSNMYLVKSSPRSFYQMM 113
TR|G3NTU3|G3NTU3_GASAC VSRAVGASLDGTERLLAACAGLQLLNTHRQDG--QVLYSNSEQARVYLTRSSPVSLYHSI 118
SP|Q92056|ASMT_CHICK    IAARLGTSIMGMERLLDACVGLKLLAVELRRE--GAFYRNTEISNIYLTKSSPKSQYHIM 107
TR|G1NPG2|G1NPG2_MELGA IAARLGTSIIGMERLLDACVGLKLLAVELRRD--GAFYRNTEISNIYLTKSSPKSQYHIM 109
TR|G3WBA1|G3WBA1_SARHA IAERLGTNSQGIEQLLDICATLKLLNVEMKKE--KALYQNTKLASLYLTKSGPKTQYNML 105
SP|D3KU67|ASMT_MUSMM    LARSSGLSPRGTRLLLDACAGLGLLRRRRGAGPRGPAYTNSPLASTFLVAGSPLSQRSLL 112
SP|D3KU66|ASMT_MOUSE    LARSSGLSPRGTRLLLDACAGLGLLRRRRGAGPRGPAYTNSPLASTFLVAGSPLSQRSLL 112
SP|B3GSH5|ASMT_RAT      VAQTGGWSPRGTQLLMDACTRLGLLR---GAG--DGSYTNSALSSTFLVSGSPQSQRCML 105
TR|E9PSX5|E9PSX5_RAT    VAQTGGWSPRGTQLLMDACTRLGLLR---GAG--DGSYTNSALSSTFLVSGSPQSQRCML 105
                         ::     .  .   *:  *. * **            : *:  :  :*.  .* :    :

SP|P46597|ASMT_HUMAN    KYMGRTSYRCWGHLADAVREGRNQYLETFGVPAEELFTAIYRSEGERLQFMQALQEVWSV 166
SP|Q8HZJ0|ASMT_MACMU    KYMGRTSYGCWGHLADAVREGKNQYLQTFGVPAEDLFKAIYRSEGERLQFMQALQEVWSV 166
SP|P10950|ASMT_BOVIN    LYAGRTAYVCWRHLAEAVREGRNQYLKAFGIPSEELFSAIYRSEDERLQFMQGLQDVWRL 166
TR|F6WEJ3|F6WEJ3_HORSE LYLARTTYHCWAHLAEAVREGRSQYLKAFGVPSEELFTAIYRSEDERLQFMRGLQDVWGV 166
TR|F6X3D5|F6X3D5_XENTR MYYSKTIYMCWHFLPDAIREGKRQYERAFGITSEDIFKDLYRSEEETVSFMHHMESIWHI 173
TR|G3NTU3|G3NTU3_GASAC QYSSRTIYLCWHYLTDAVREGRNQYEKAFGVSSNDLFQALYRSDEEMVKFMHLMNSIWNI 178
SP|Q92056|ASMT_CHICK    MYYSNTVYLCWHYLTDAVREGRNQYERAFGISSKDLFGARYRSEEEMLKFLAGQNSIWSI 167
TR|G1NPG2|G1NPG2_MELGA MYYSNTVYLCWHYLTDAVREGRNQYERAFGISSEDLFRAMYRSEEEMLKFLAGQNSIWSI 169
TR|G3WBA1|G3WBA1_SARHA LYMSKTTYSCWNHLTEAIREGKNQYMRTFGIPSDDLFEAIYRSEEELIKFMHGLHEIWSI 165
SP|D3KU67|ASMT_MUSMM    LYLAGTTYLCWGHLADGVREGRSQYARAVGVDADDPFTAIYRSEAERLLFMRGLQETWSL 172
SP|D3KU66|ASMT_MOUSE    LYLAGTTYLCWGHLADGVREGRSQYARAVGVDADDPFTAIYRSEAERLLFMRGLQETWSL 172
SP|B3GSH5|ASMT_RAT      LYLAGTTYGCWAHLAAGVREGRNQYSRAVGISAEDPFSAIYRSEPERLLFMRGLQETWSL 165
TR|E9PSX5|E9PSX5_RAT    LYLAGTTYGCWAHLAAGVREGRNQYSRAVGISAEDPFSAIYRSEPERLLFMRGLQETWSL 165
                          * . * * ** .*. .:***: ** .:.*: :.: *   ***: * : *:   .. * :

SP|P46597|ASMT_HUMAN    NGRSVLTAFDLSVFPLMCDLGGGAGALAKECMSLYPGCKITVFDIPEVVWTAKQHFSFQ- 225
SP|Q8HZJ0|ASMT_MACMU    NGRSVLTAFDLSGFPLMCDLGGGPGALAKECLSLYPGCKVTVFDVPEVVRTAKQHFSFP- 225
SP|P10950|ASMT_BOVIN    EGATVLAAFDLSPFPLICDLGGGSGALAKACVSLYPGCRAIVFDIPGVVQIAKRHFSAS- 225
TR|F6WEJ3|F6WEJ3_HORSE DGRSVLSAFDLSSFPVICDLGGCSGALAKECVSLYPGCRVTVFDIPEVVQTAKTHFPLP- 225
TR|F6X3D5|F6X3D5_XENTR CGKDVLAAFDLSSFKEICDIGGCSGGLAKHFLSLYPSSSVTIMDLPEVVQMAKKHFITD- 232
TR|G3NTU3|G3NTU3_GASAC CGKDVVTAFDLSPFNVICDLGGCSGALAKQCTSAYPECTVTIFDLPKVVQTSREHFVTD- 237
SP|Q92056|ASMT_CHICK    CGRDVLTAFDLSPFTQIYDLGGGGGALAQECVFLYPNCTVTIYDLPKVVQVAKERLVPP- 226
TR|G1NPG2|G1NPG2_MELGA CGRDVLTAFDLSPFTQIYDLGGGGGALAQECVFLYPNCTVTIYDLPKVVQVAKERLVPP- 228
TR|G3WBA1|G3WBA1_SARHA NGRYVITAFDLSSFPLIYDLGGCSGALAKECITIYPECQVTIFEKEEVVQVAKKHFSFL- 224
SP|D3KU67|ASMT_MUSMM    CGGRVLAAFDLSPFRVICDLGGGSGALARMAARLYPGSEVTVFETPDVVAAARAHFPPPA 232
SP|D3KU66|ASMT_MOUSE    CGGRVLTAFDLSPFRVICDLGGGSGALARMAARLYPGSEVTVFETPDVVAAARAHFPPPA 232
SP|B3GSH5|ASMT_RAT      CGGRVLTAFDLSRFRVICDLGGGSGALAQEAARLYPGSSVCVFDLPDVIAAARTHFLSP- 224
TR|E9PSX5|E9PSX5_RAT    CGGRVLTAFDLSRFRVICDLGGGSGALAQEAARLYPGSSVCVFDLPDVIAAARTHFLSP- 224
                          *  *::***** *  : *:**  *.**:     ** .   : :   *:  :: ::
```

Fig. 5. Clustal alignment of the ASMT protein from different species. Residues highlighted in green, yellow or red display normal, reduced or disrupted ASMT activity, respectively.

```
SP|P46597|ASMT_HUMAN    ----EEEQIDFQEGDFFKDPLPEADLYILARVLHDWADGKCSHLLERIYHTCKP----GG 277
SP|Q8HZJ0|ASMT_MACMU    ----EEEEIHLQEGDFFKDPLPEADLYILARILHDWADGKCSHLLERVYHTCKP----GG 277
SP|P10950|ASMT_BOVIN    ----EDERISFHEGDFFKDALPEADLYILARVLHDWTDAKCSHLLQRVYRACRT----GG 277
TR|F6WEJ3|F6WEJ3_HORSE ----EGARIRFCEGDFFKDPLPEADLYILARILHDWTDEKCAHLLERVYRACKP----GG 277
TR|F6X3D5|F6X3D5_XENTR ----GD-IVFLEEGDFFNDPLPESDLYILARIIHDWTEDKCLRLLNKIYKSCRPGIKTGG 287
TR|G3NTU3|G3NTU3_GASAC ----ADGRINFYEGDFFKDPLPEADLYILARILHDWTDERCTELLSRVNRDVLT-CRTRG 292
SP|Q92056|ASMT_CHICK    ----EERRIAFHEGDFFKDSIPEADLYILSKILHDWDDKKCRQLLAEVYKACRP----GG 278
TR|G1NPG2|G1NPG2_MELGA ----EERRIAFHEGDFFKDSIPEADLYILSKILHDWDDEKCGQLLAEVYKACRP----GG 280
TR|G3WBA1|G3WBA1_SARHA ----ENPRIRFHIGDFFRDPIPEADLYILARILHDWADNKCIQLLKNIHHACRP----GC 276
SP|D3KU67|ASMT_MUSMM    DEDGAEPRVRFLSGDFFRSPLPPADLYVLARVLHDWADAACVELLRRVRGALRP----GG 288
SP|D3KU66|ASMT_MOUSE    DEDGAEPRVRFLSGDFFRSPLPPADLYVLARVLHDWADAACVELLRRVRGALRP----GG 288
SP|B3GSH5|ASMT_RAT      ---GARPSVRFVAGDFFRSRLPRADLFILARVLHDWADGACVELLGRLHRACRP----GG 277
TR|E9PSX5|E9PSX5_RAT    ---GARPSVRFVAGDFFRSRLPRADLFILARVLHDWADGACVELLGRLHRACRP----GG 277
                                : :  ****.. :* :**::*::::*** :  * .** .:     .

SP|P46597|ASMT_HUMAN    GILVIESLLDEDRRGPLLTQLYSLNMLVQTEGQERTPTHYHMLLSSAGFRDFQFKKTGAI 337
SP|Q8HZJ0|ASMT_MACMU    GILVIESLLDEDRRGPLLTQLYSLNMLVQTEGQERTPTHYHMLLSSAGFRDFQFKKTGAI 337
SP|P10950|ASMT_BOVIN    GILVIESLLDTDGRGPLTTLLYSLNMLVQTEGRERTPAEYRALLGPAGFRDVRCRRTGGT 337
TR|F6WEJ3|F6WEJ3_HORSE GVLVVEGLLDADGRGPLTAQLWSLNMLVQTEGRERTPAQYCELLGAAGFGTVCCRRTRGL 337
TR|F6X3D5|F6X3D5_XENTR GVLLVEALLNEDRSGPLSSQMFSLNMLLQTEGKERSASEYHKLLADSGFREIQVKATGKF 347
TR|G3NTU3|G3NTU3_GASAC GVLLVEALLHEDGSGPLTVQLYSLNMLVQTEGRERTAAQYAALLAAAGFTAVRHRLTGKI 352
SP|Q92056|ASMT_CHICK    GVLLVESLLSEDRSGPVETQLYSLNMLVQTEGKERTAVEYSELLGAAGFREVQVRRTGKL 338
TR|G1NPG2|G1NPG2_MELGA GVLLVESLLSGDRSGPVETQLYSLNMLVQTEGKERTAAEYSKLLGAAGFRDIRVRRTGKL 340
TR|G3WBA1|G3WBA1_SARHA GVLVVETLLAEDKRGPLTSQLYSLNMLVQTEGKERTPTEYHKILTQAGFQNFQFKKTGKI 336
SP|D3KU67|ASMT_MUSMM    AVLLVESVLSPGGAGPTRTLLLSLTMLLQARGRERTEAEYRALTARAGFSRLRLRRPRGP 348
SP|D3KU66|ASMT_MOUSE    AVLLVESVLSPGGAGPTRTLLLSLTMLLQARGRERTEAEYRALTARAGFSRLRLRRPRGP 348
SP|B3GSH5|ASMT_RAT      ALLLVEAVLAKGGAGPLRSLLLSLNMMLQAEGWERQASDYRNLATRAGFPRLQLRRPGGP 337
TR|E9PSX5|E9PSX5_RAT    -VPLVEAVLAKGGAGPLRSLLLSLNMMLQAEGWERQASDYRNLATRAGFPRLQLRRPGGP 336
                         : ::* :*  .  **    : **.*::*:.* **   .*  :   :**  .  : .

SP|P46597|ASMT_HUMAN    YDAILARK---------------------------------------------------- 345
SP|Q8HZJ0|ASMT_MACMU    YDAILVRK---------------------------------------------------- 345
SP|P10950|ASMT_BOVIN    YDAVLARK---------------------------------------------------- 345
TR|F6WEJ3|F6WEJ3_HORSE YHAVLATR---------------------------------------------------- 345
TR|F6X3D5|F6X3D5_XENTR YDAVLGKK---------------------------------------------------- 355
TR|G3NTU3|G3NTU3_GASAC YDAVLGRKEG--------------------------------------------------- 362
SP|Q92056|ASMT_CHICK    YDAVLGRK---------------------------------------------------- 346
TR|G1NPG2|G1NPG2_MELGA YDAVLGRK---------------------------------------------------- 348
TR|G3WBA1|G3WBA1_SARHA YDAILARK---------------------------------------------------- 344
SP|D3KU67|ASMT_MUSMM    YHAMMAARG-------------------------------GGAGARS-DGGGGEATSQT 375
SP|D3KU66|ASMT_MOUSE    YHAMMAARG-------------------------------GGAGARS-DGGGGDATSQT 375
SP|B3GSH5|ASMT_RAT      YHAMLARRGPRPGIITGVGSNTTGTGSFVTGIRRDVPGARSDAAGTGSGTGNTGSGIMLQ 397
TR|E9PSX5|E9PSX5_RAT    YHAMLAR---------------------------------DPALGS-----SQEWEVTLR 358
                         *.*::

SP|P46597|ASMT_HUMAN    ------------------------------------
SP|Q8HZJ0|ASMT_MACMU    ------------------------------------
SP|P10950|ASMT_BOVIN    ------------------------------------
TR|F6WEJ3|F6WEJ3_HORSE ------------------------------------
TR|F6X3D5|F6X3D5_XENTR ------------------------------------
TR|G3NTU3|G3NTU3_GASAC ------------------------------------
SP|Q92056|ASMT_CHICK    ------------------------------------
TR|G1NPG2|G1NPG2_MELGA ------------------------------------
TR|G3WBA1|G3WBA1_SARHA ------------------------------------
SP|D3KU67|ASMT_MUSMM    GSGTGREVGAQD----------------------- 387
SP|D3KU66|ASMT_MOUSE    GSGTGSEVGAQD----------------------- 387
SP|B3GSH5|ASMT_RAT      GETLESEVSAPQAGSDVGGAGNEPRSGTLKQGDWK 432
TR|E9PSX5|E9PSX5_RAT    GQEVLSQV--------------------------- 366
```

Fig. 5. (Continued)

(Fig. 5). Around the SAM binding site, most residues are strictly conserved, especially the aromatic residues F143, Y147, F156, F237, F238, W257, as well as G187, G235 and D236. Other important residues such as D210 and R252 show small variations in size, but not in charge or polarity, so that the H-bonds to the SAM moiety are maintained (see Fig. 3). At the NAS binding site, residues are more strictly conserved near the hydroxyl side of NAS, but more variable at the acetamide end. These strictly conserved residues include Y108, F156, M303, Q306, and Y338. For those residues, which bind the acetamide group, N302 is a threonine in some cases, whereas T307 is sometimes an alanine. As expected, the presumed active residues (H255, D256, and E311), which are responsible for methyl transfer and proton abstraction, are strictly conserved. Elsewhere, much of the beta sheet of the C-terminal (SAM-binding) domain also is highly conserved, and many residues of the dimer interface are also well conserved among ASMT sequences. This latter group occurs along the helices of the central 4-helix bundle, and it hints at a possible functional role for the dimer interface. Finally, the only other zone of strictly conserved residues is the loop V171-F180, and intriguingly, this segment is well removed from the active site and far from the dimer interface.

The human *ASMT* gene reveals three possible isoforms, which result from alternative splicings of exon 6 and 7 [33, 34]. The structure here is of the first and major isoform, P46597-1, which corresponds to the active ASMT protein identified in other species. In isoform P46597-2, a segment of 47 residues (G188-F237), encoded by the exon 7, is deleted with respect to the P46597-1 sequence, and this segment (two helices and two beta strands) corresponds to one side of the SAM O-MT domain that lines the SAM binding site (Fig. 6). Thus, P46597-2 is not expected to bind SAM tightly, nor show any O-MT activity (Fig. 1C). Interestingly, the points of excision (residues 188 and 236) lie within proximity of each other (12.4 Å) and are close enough to be relinked together upon a simple rearrangement. Isoform P46597-3, however, has an insertion of 28 residues between G187 and G188, just upstream of the isoform P46597-2 deletion. This additional sequence is encoded by exon 6 and displays homology with a Long INterspread Element (LINE) [11]. The location of the insertion of isoform P46597-3 occurs just after G187, which is a highly conserved and buried residue that makes a H-bond to the SAM molecule. Consequently, this insertion is expected to distort this region and destroy the ability to bind SAM, Indeed, biochemical studies show that no ASMT activity is detected for isoform P46597-3 (Fig. 1C).

The genetic variability of *ASMT* was studied through the 1000 genomes project [35], as well as through specific *ASMT* mutation screening [7–11]. Overall, 27 nonsynonymous variants and 1 splice mutation were identified (Tables 2 and 3). Rare variants affecting the protein sequence of ASMT were reported in patients with neuropsychiatric disorders such as ASD, ADHD, ID, and bipolar disorder [7–11]. ASMT variants were also found in the general population including relatively frequent polymorphisms in African (K81E, rs141964565) and Han Chinese (N17K, rs17149149) populations [7, 9, 11]. To test whether these variants affect the enzyme activity, we transfected wild-type or mutant human *ASMT* cDNA in COS cells and ascertained their enzyme activities using radioenzymology. We tested 20 nonsynonymous variants, which were confirmed by Sanger sequence technology [7–11], and we found that the majority of these mutations reduced or abolished ASMT activity (Table 3 and Fig. 1C). Notably, the PolyPhen2 algorithm was correct for predicting the functional impact of almost all mutations (Table 3). Variants that displayed $< 10\%$ of ASMT activity either form part of the SAM binding site or are near the substrate binding site (Fig. 1D). For example, the D210G variant removes the carboxylic acid group, which forms two direct H-bonds to the sugar hydroxyl groups of the SAM moiety. Such a structural loss would clearly disrupt SAM binding. For the N17K variant, which is near the NAS methyl group, it would add a positive charge among a predominantly hydrophobic environment and distort the shape of the NAS binding site. Similarly, for the T296M variant, where the change in its size and H-bonding character would perturb the contacts that anchor an alpha helix packing against the NAS molecule. Some variants are located at the dimerization interface (N17K, R291Q & L298F) implying that the dimer is critical for the formation the NAS binding sites. Changes in the size (L298F) and charge (N17K) presumably either perturb the folding of the polypeptide chain or deform the binding pocket sufficiently to alter the orientation of the NAS molecule, thus moving it away from the activated methyl group of the SAM moiety. Likewise, a buried variant, Y248H, which could be considered to be relatively far from the SAM binding site (12Å) shows $< 10\%$ activity (Table 3 and Fig. 1C). Coincidentally, the phenol oxygen of this tyrosine forms an H-bond with a carboxyl group of a proline (P243), which in turn is also a variant found in a patient with bipolar disorder (P243L, 10–20% activity). Although this proline is 10 Å away from the SAM molecule, it is part of a long loop (232–247), which forms one side of the SAM binding site. Presumably, such variants located in the secondary coordination sphere of the SAM binding site are critical for activity. Furthermore, it should be noted that this entire region (235–257) is highly conserved among all ASMTs sequenced and contains the residues H255 and D256, which are at the reaction center for the methyl transfer. Finally, the ASMT crystal structures cannot provide a clear explanation of loss of activity for certain variants. The reasons for the loss of activity in the variant R291Q are obscure because this residue is far (> 15 Å) from the NAS binding site and it is located on the surface in a loop, which forms part of the dimer interface. R291 is not well conserved among the 13 ASMT sequences compared, where the corresponding residue may be serine or alanine.

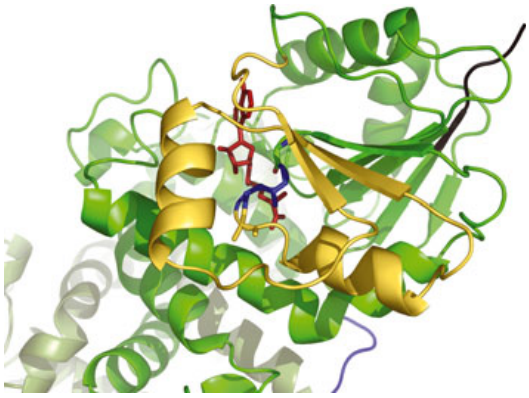


Fig. 6. Ribbon diagram of acetyl serotonin methyltransferase (ASMT) showing the locations of the three isoforms. Isoform-1 is the crystal structure presented here (green, blue, and yellow). In yellow is the fragment deleted in isoform P46597-2 and in blue is the location of the insert for P46597-3. The molecule of *S*-adenosyl-L-methionine (SAM) is represented in red.

Discussion

Methyltransferases are present in a wide range of species and are implicated in many processes of life. Despite the number of structures available of O-MTs (over 700 entries in the PDB), there is no representative example for ASMT. Here, we have determined the crystal structures of human ASMT in complex with its substrate and cofactor. As expected, the overall fold of the SAM binding site is strictly conserved, and the major differences with other SAM

Table 2. Acetyl serotonin methyltransferase (ASMT) variations identified during the 1000 genomes project and listed in dbSNP.

						Allele frequency in the populations of the 1000 genomes project (%)				
Variation	ID	Chr: bp	Alleles	Source	PolyPhen2	Africa	Asia	America	Europe	Total
R10C	1KG_X_1734120	X:1734120	C/T	1KG	Probably damaging	0	0.17	0	0	0.046
Y15S	1KG_X_1734136	X:1734136	A/C	1KG	Probably damaging	0	0.35	0	0	0.091
N17K	rs17149149	X:1734143	C/A	dbSNP	Possibly damaging	0.20	2.45[a]	0.55	0	0.78
V46M	1KG_X_1742098	X:1742098	G/A	1KG	Benign	1.63	0	0	0.13	0.41
K81E	1KG_X_1742203	X:1742203	A/G	1KG	Benign	3.45	0	0.28	0	0.82
G151S	1KG_X_1748721	X:1748721	G/A	1KG	Benign	0	0.69	0	0	0.18
A174T	1KG_X_1748790	X:1748790	G/A	1KG	Possibly damaging	0	0	0.83	0	0.14
P202S	1KG_X_1752084	X:1752084	C/T	1KG	Probably damaging	0	0	0	0.13	0.046
I269M	1KG_X_1755434	X:1755434	C/G	1KG	Possibly damaging	0.81	0	0	0	0.18
L298F	rs121918822	X:1761761	C/T	1KG	Probably damaging	0	0	0	0.39	0.14
T307M	1KG_X_1761789	X:1761789	C/T	1KG	Probably damaging	0.20	0	0.28	0	0.091
IVS5 + 2T > C	1KG_X_1748834	X:1748834	T/C	1KG	Damaging[b]	0	0	0	0.13	0.046
Benign						5.08	0.69	0.28	0.13	1.42
Damaging						1.22	2.97	1.66	0.66	1.55
All						6.30	3.67	1.93	0.79	2.97

The numbers of individuals in each population are Africa (N = 246), Asia (N = 286), America (N = 181), and Europe (N = 381).
[a]The allelic frequency of rs17149149 (N17K) is 6.7% in the 45 unrelated Han Chinese in Beijing, China, from the International HapMap project.
[b]The variation IVS5 + 2T > C is located at the acceptor splice site of exon 5 and was shown to alter the splicing of *ASMT* (Melke et al.[11]).

methyltransferases occur at the second substrate binding site, which defines the specificity for NAS of these enzymes. To better understand the catalytic mechanism, we attempted to solve the structure of this enzyme in its apo-form and also in complex with the cofactor SAM and substrate NAS both soaked in excess. However, to our surprise, we found electron density for a SAM molecule in the data set collected from supposedly apo-enzyme crystals. Presumably, the recombinant protein, purified from *Leishmania tarentolae,* acquired the cosubstrate SAM from the cytosol. Consequently, the absence of the structure of the apo-form deprives us the possibility of interpreting any of conformational changes following the binding of the cofactor SAM.

When compared to other SAM-dependant O-MTs of the same family, the structures presented here are in a tightly closed conformation, which brings together the SAM and the substrate to form the enzyme–substrates complex. It shows all the residues implicated in binding and catalysis. The majority of these residues are conserved in all ASMT proteins identified to date. The comparison of the positions of the amino acids interacting with the methyl group of the SAM and the hydroxyl group of the indole of NAS reveal that H255 and E311 are likely to play key roles in the methyl transfer. The same corresponding residues are also proposed to be involved for others O-MTs. Via excess soaking experiments in crystals, we succeeded in binding a NAS molecule into crystals containing SAM. Examination of the two crystal structures reveals that the addition of NAS reduces the thermal parameters of two segments of the polypeptide chain (Y131-F156 and P212-D236) and the SAM molecule with respect to the rest of the protein (Fig. 6). The fact that NAS stiffens these parts of the polypeptide chain suggests conformational changes, which could expose the SAM and NAS binding sites to the solvent. One of these segments, P212-D236, resides next to the SAM binding site, and it corresponds to the C-terminal domain movement seen between open and closed conformations of other O-MTs [29]. The other segment, Y131-F156, covers the NAS binding site, and its movement could provide an access to this site. Thus, the sequential binding mechanism observed by others could be simply explained by a less buried and thus more solvent-accessible SAM binding site compared with the deeply buried NAS binding site.

With the crystal structures of ASMT presented here, we can offer explanations for the role of virtually all mutations inducing the total loss of the enzymatic activity (< 10% of wild-type activity). The majority of the mutations occur at residues located near the substrate binding sites or in the dimerization domain. Moreover, we verified the position and the nature of these residues in all ASMTs sequenced to date and found that they are strictly conserved, thus implying that they play an important role in preserving the native conformational structure and the enzyme activity of ASMT. For the mutations abolishing around 50% of the activity, it is not always easy to pinpoint their possible roles in the enzymatic activity just by examining their positions in the 3-dimensional structure of the protein. We only know that the majority of these mutations are localized in the secondary sphere of the substrate binding site, and when we analyze their position in the sequence of other ASMTs, they are less conserved; this is more obvious when the mutation does not affect the activity.

A certain amount of caution should be employed when interpreting the structure–function relationships of any variant as any sequence mutation can have unexpected changes to the polypeptide fold. Furthermore, as proteins interact with one another, surface residues, which might be expected to show little influence on the activity, could be critical for the binding of physiological partners. In the

Table 3. Enzyme activity and prediction of the functional impact of the human acetyl serotonin methyltransferase (ASMT) variants identified in patients and in the general population

ASMT variants[a]	ASMT activity[b] (% of WT activity)	Polyphen2 (Div/Var)	nsSNP Analyzer	PMUT	SNAP	SIFT	PANTHER	Populations
Normal activity (>80%)								
N13H	98 ± 16	Benign/Benign	Unknown	Neutral	Neutral	Tolerated	Tolerated	ID, GP
K81E	98 ± 27	Benign/Benign	Unknown	Neutral	Neutral	Tolerated	Deleterious	ASD, GP
R111K	106 ± 12	Benign/Benign	Unknown	Neutral	Neutral	Tolerated	Deleterious	BP
K219R	84 ± 6	Benign/Benign	Unknown	Neutral	Neutral	Tolerated	Deleterious	GP
E288D	86 ± 9	Benign/Benign	Neutral	Neutral	Neutral	Tolerated	Deleterious	ASD, BP, ID, GP
Reduced activity (10–80%)								
L11F	50 ± 8	Poss. dam./Benign	Unknown	Neutral	Neutral	Tolerated	Tolerated	ASD
E61Q	55 ± 1	Poss. dam./Benign	Unknown	Neutral	Neutral	Tolerated	Deleterious	GP
I269M	56 ± 0.7	Poss. dam./Poss. dam.	Unknown	Neutral	Non-Neutral	Affect protein function	Deleterious	ASD
P243L	25 ± 2	Prob. dam./Prob. dam.	Disease	Pathological	Non-Neutral	Affect protein function	Deleterious	BP, GP
C273S	44 ± 3	Prob. dam./Prob. dam.	Unknown	Neutral	Non-Neutral	Affect protein function	Deleterious	ASD, GP
G278A	52 ± 9	Poss. dam./Poss. dam.	Unknown	Neutral	Neutral	Tolerated	Deleterious	ASD
V305M	39 ± 2	Prob. dam./Prob. dam.	Unknown	Neutral	Neutral	Affect protein function	Deleterious	BP
H318D	59 ± 7	Benign/Benign	Unknown	Pathological	Neutral	Tolerated	Deleterious	ASD
Disrupted activity (<10%)								
N17K[c]	4 ± 1	Poss. dam./Poss. dam.	Unknown	Neutral	Non-Neutral	Tolerated	Deleterious	ASD, ID, GP
V171M	4 ± 1	Prob. dam./Prob. dam.	Unknown	Neutral	Neutral	Affect protein function	Deleterious	ID
D210G	4 ± 1	Prob. dam./Prob. dam.	Disease	Pathological	Non-Neutral	Affect protein function	Deleterious	ADHD, GP
Y248H	8 ± 3	Prob. dam./Prob. dam.	Unknown	Neutral	Non-Neutral	Affect protein function	Deleterious	BP
R291Q	5 ± 1	Prob. dam./Poss. dam.	Unknown	Pathological	Neutral	Tolerated	Deleterious	GP
T296M	4 ± 1	Poss. dam./Poss. dam.	Unknown	Pathological	Neutral	Affect protein function	Deleterious	ID
L298F	6 ± 2	Poss. dam./Poss. dam.	Disease	Neutral	Neutral	Affect protein function	Deleterious	ASD, ADHD, ID, BP, GP

[a]The nomenclature of *ASMT* variations is based on the sequence P46597-1, the protein sequence coding for the functional isoform of ASMT, and therefore is different from that in the paper by Melke et al. [11] L326F is now L298F.
[b] The wild-type ASMT enzyme activity was 21 ± 2 nmol/h/mg prot.
[c]N17K variant is rs17149149 and was found at the heterozygote state in 13.3% of the Han Chinese population from the HAPMAP consortium.
ASD, autism spectrum disorders; ADHD, attention-deficit hyperactivity disorders; BP, bipolar disorder; ID, intellectual disability; GP, general population.

context of melatonin biosynthesis, AANAT, the first enzyme of the pathway, binds to the scaffold protein 14-3-3 (PDB ID: 1ib1) [36], and the resulting structural changes in AANAT bring about an increase in the affinity for serotonin, which is converted to NAS. In pinealocytes, Maronde et al. [13] discovered that AANAT, 14-3-3, and ASMT colocalize along fibril-like structures called pinealocyte processes, into which melatonin is secreted. The colocalization of these enzymes in the melatonin biosynthetic pathway is strong evidence for the existence of a multiprotein complex dedicated to melatonin synthesis, which the authors have named the 'melatoninosome'. In light of these results, it is more than likely that certain surface residues of the ASMT dimer could bind AANAT, 14-3-3, or other proteins as part of a larger assembly dedicated to melatonin synthesis. Thus, certain variants of ASMT may have profound effects upon activity because they are implicated in the binding to physiologically important protein partner rather than their proximity to the active site.

We have shown that nonsynonymous variants identified in human populations could affect crucial amino acids required for ASMT activity and therefore most likely contribute to melatonin deficiency in vivo [11]. Based on the 1000 genome project, the allelic frequency of *ASMT* deleterious mutations in human populations ranges from 0.66% in Europe to 2.97% in Asia (Table 2). In particular, the deleterious N17K variant was found in relatively high allelic frequency (2.45%) in 200 Han Chinese individuals of the 1000 genomes project [35]. This frequency was even higher (6.7%) in the 90 independent Han Chinese subjects from the International HapMap project [37]. Taken together, this variant is present in almost 9% of the Han Chinese (allelic frequency: 4.48%). The presence of a single deleterious *ASMT* allele might not be dramatically harmful, but a low level of enzyme activity in combination

with additional genetic/epigenetic and environmental susceptibility factors might increase the individual risk of melatonin deficiency. Sequencing the entire exome/genome sequence of patients is now available, and such approaches will reveal novel *ASMT* variations. In line with this, the crystal structure of the human ASMT enzyme and the mapping of functionally deleterious variants will be essential for the interpretation of the genomic makeup of patients with melatonin deficiency and important for their diagnosis.

Acknowledgements

This work was supported by the Institut Pasteur, University Paris Diderot, CNRS, Fondation Orange, ANR (ANR-08-MNPS-037-01 – SynGen), Neuron-ERANET (EUHF-AUTISM), Fondation FondaMentale, Fondation Recherche Médicale, the ESRF and Synchrotrons (SOLEIL Proposal 20080516, ESRF at Grenoble).

Contributions

HGB, PL, AH, WS, and TB designed, carried out, and analyzed the experiments and wrote the manuscript. CP, GH, VB, PW, MBA, NL, JB and PB designed and carried out the experiments. EM advised and assisted on all aspects of the project and wrote the manuscript.

References

1. Reiter RJ, Tan DX, Fuentes-Broto L. Melatonin: a multitasking molecule. Prog Brain Res 2010; **181**:127–151.
2. Hallam KT, Olver JS, Chambers V et al. The heritability of melatonin secretion and sensitivity to bright nocturnal light in twins. Psychoneuroendocrinology 2006; **31**:867–875.
3. Stehle JH, Saade A, Rawashdeh O et al. A survey of molecular details in the human pineal gland in the light of phylogeny, structure, function and chronobiological diseases. J Pineal Res 2011; **51**:17–43.
4. Barnard AR, Nolan PM. When clocks go bad: neurobehavioural consequences of disrupted circadian timing. PLoS Genet 2008; **4**:e1000040.
5. Shimomura K, Lowrey PL, Vitaterna MH et al. Genetic suppression of the circadian Clock mutation by the melatonin biosynthesis pathway. Proc Natl Acad Sci U S A 2010; **107**:8399–8403.
6. Simonneaux V, Ribelayga C. Generation of the melatonin endocrine message in mammals: a review of the complex regulation of melatonin synthesis by norepinephrine, peptides, and other pineal transmitters. Pharmacol Rev 2003; **55**:325–395.
7. Chaste P, Clement N, Botros HG et al. Genetic variations of the melatonin pathway in patients with attention-deficit and hyperactivity disorders. J Pineal Res 2011; **51**:394–399.
8. Jonsson L, Ljunggren E, Bremer A et al. Mutation screening of melatonin-related genes in patients with autism spectrum disorders. BMC Med Genomics 2010; **3**:10.
9. Pagan C, Botros HG, Poirier K et al. Mutation screening of ASMT, the last enzyme of the melatonin pathway, in a large sample of patients with intellectual disability. BMC Med Genet 2011; **12**:17.
10. Toma C, Rossi M, Sousa I et al. Is ASMT a susceptibility gene for autism spectrum disorders? A replication study in European populations. Mol Psychiatry 2007; **12**:977–979.
11. Melke J, Goubran Botros H, Chaste P et al. Abnormal melatonin synthesis in autism spectrum disorders. Mol Psychiatry 2008; **13**:90–98.
12. Ben-Abdallah M, Bondet V, Fauchereau F et al. Production of soluble, active acetyl serotonin methyl transferase in Leishmania tarentolae. Protein Expr Purif 2010; **75**:114–118.
13. Maronde E, Saade A, Ackermann K et al. Dynamics in enzymatic protein complexes offer a novel principle for the regulation of melatonin synthesis in the human pineal gland. J Pineal Res 2011; **51**:145–155.
14. Santarsiero BD, Yegian DT, Lee CC et al. An approach to rapid protein crystallization using nanodroplets. J Appl Cryst 2002; **35**:278–281.
15. Kabsch W. Xds. Acta Crystallogr D Biol Crystallogr 2010; **66**:125–132.
16. Schneider TR, Sheldrick GM. Substructure solution with SHELXD. Acta Crystallogr D Biol Crystallogr 2002; **58**:1772–1779.
17. Mccoy AJ. Solving structures of protein complexes by molecular replacement with Phaser. Acta Crystallogr D Biol Crystallogr 2007; **63**:32–41.
18. Cowtan K. Recent developments in classical density modification. Acta Crystallogr D Biol Crystallogr 2010; **66**:470–478.
19. Cowtan K. The Buccaneer software for automated model building. 1. Tracing protein chains. Acta Crystallogr D Biol Crystallogr 2006; **62**:1002–1011.
20. Vagin A, Teplyakov A. Molecular replacement with MOLREP. Acta Crystallogr D Biol Crystallogr 2010; **66**:22–25.
21. Blanc E, Roversi P, Vonrhein C et al. Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. Acta Crystallogr D Biol Crystallogr 2004; **60**:2210–2221.
22. Emsley P, Lohkamp B, Scott WG et al. Features and development of Coot. Acta Crystallogr D Biol Crystallogr 2010; **66**:486–501.
23. Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. J Mol Biol 2007; **372**:774–797.
24. Axelrod J, Wurtman RJ, Snyder SH. Control of hydroxyindole O-methyltransferase activity in the rat pineal gland by environmental lighting. J Biol Chem 1965; **240**:949–954.
25. Bernard M, Donohue SJ, Klein DC. Human hydroxyindole-O-methyltransferase in pineal gland, retina and Y79 retinoblastoma cells. Brain Res 1995; **696**:37–48.
26. Martin JL, Mcmillan FM. SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold. Curr Opin Struct Biol 2002; **12**:783–793.
27. Zubieta C, He XZ, Dixon RA et al. Structures of two natural product methyltransferases reveal the basis for substrate specificity in plant O-methyltransferases. Nat Struct Biol 2001; **8**:271–279.
28. Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr D Biol Crystallogr 2004; **60**:2256–2268.
29. Liu CJ, Deavours BE, Richard SB et al. Structural basis for dual functionality of isoflavonoid O-methyltransferases in the evolution of plant defense responses. Plant Cell 2006; **18**:3656–3669.
30. Zubieta C, Kota P, Ferrer JL et al. Structural basis for the modulation of lignin monomer methylation by caffeic acid/

5-hydroxyferulic acid 3/5-O-methyltransferase. Plant Cell 2002; **14**:1265–1277.
31. Vidgren J, Svensson LA, Liljas A. Crystal structure of catechol O-methyltransferase. Nature 1994; **368**:354–358.
32. Cui FC, Pan XL, Liu W et al. Elucidation of the methyl transfer mechanism catalyzed by chalcone O-methyltransferase: a density functional study. J Comput Chem 2011; **32**:3068–3074.
33. Rodriguez IR, Mazuruk K, Schoen TJ et al. Structural analysis of the human hydroxyindole-O-methyltransferase gene. Presence of two distinct promoters. J Biol Chem 1994; **269**:31969–31977.
34. Donohue SJ, Roseboom PH, Illnerova H et al. Human hydroxyindole-O-methyltransferase: presence of LINE-1 fragment in a cDNA clone and pineal mRNA. DNA Cell Biol 1993; **12**:715–727.
35. Durbin RM, Abecasis GR, Altshuler DL et al. A map of human genome variation from population-scale sequencing. Nature 2010; **467**:1061–1073.
36. Obsil T, Ghirlando R, Klein DC et al. Crystal structure of the 14-3-3zeta:serotonin *N*-acetyltransferase complex. a role for scaffolding in enzyme regulation. Cell 2001; **105**:257–267.
37. Hapmap Consortium. The International HapMap Project. Nature 2003; **426**:789–796.

J. Pineal Res. 2011; 51:394–399
Doi:10.1111/j.1600-079X.2011.00902.x

Journal of Pineal Research

Journal of Pineal Research
Molecular, Biological, Physiological and Clinical Aspects of Melatonin

Genetic variations of the melatonin pathway in patients with attention-deficit and hyperactivity disorders

Abstract: Melatonin is a powerful antioxidant and a synchronizer of many physiological processes. Alteration in melatonin signaling has been reported in a broad range of diseases, but little is known about the genetic variability of this pathway in humans. Here, we sequenced all the genes of the melatonin pathway – *AA-NAT*, *ASMT*, *MTNR1A*, *MTNR1B* and *GPR50* – in 321 individuals from Sweden including 101 patients with attention-deficit/hyperactivity disorder (ADHD) and 220 controls from the general population. We could find several damaging mutations in patients with ADHD, but no significant enrichment compared with the general population. Among these variations, we found a splice site mutation in *ASMT* (IVS5 + 2T > C) and one stop mutation in *MTNR1A* (Y170X) – detected exclusively in patients with ADHD – for which biochemical analyses indicated that they abolish the activity of ASMT and MTNR1A. These genetic and functional results represent the first comprehensive ascertainment of melatonin signaling deficiency in ADHD.

Pauline Chaste[1,2]*, Nathalie Clement[3,4,]*, Hany Goubran Botros[1,2], Jean-Luc Guillaume[3,4], Marina Konyukh[1,2], Cécile Pagan[1,2], Isabelle Scheid[1,2], Gudrun Nygren[5], Henrik Anckarsäter[5,6], Maria Rastam[7], Ola Ståhlberg[5], I. Carina Gillberg[5], Jonas Melke[8], Richard Delorme[1,2,9], Claire Leblond[1,2], Roberto Toro[1,2], Guillaume Huguet[1,2], Fabien Fauchereau[1,2,10], Christelle Durand[1], Lydia Boudarene[1], Emilie Serrano[1], Nathalie Lemière[1,2], Jean Marie Launay[11], Marion Leboyer[12,13], Ralf Jockers[3,4], Christopher Gillberg[5,14] and Thomas Bourgeron[1,2,10]

[1]Human Genetics and Cognitive Functions, Institut Pasteur, Paris, France; [2]Genes, Synapses and Cognition, CNRS URA 2182, Institut Pasteur, Paris, France; [3]Institut Cochin, University Paris Descartes, CNRS (UMR 8104), Paris, France; [4]Inserm U567, Paris, France; [5]Department of Child and Adolescent Psychiatry, Göteborg University, Göteborg, Sweden; [6]Institute of Clinical Sciences, Lund University, Malmö, Sweden; [7]Department of Clinical Sciences in Lund, Lund University, Lund, Sweden; [8]Department of Pharmacology, Institute of Neuroscience and Physiology, Gothenburg University, Göteborg, Sweden; [9]Department of Child and Adolescent Psychiatry, Robert Debré Hospital, Assistance Publique-Hôpitaux de Paris, Paris, France; [10]University Denis Diderot Paris 7, Paris, France; [11]Department of Biochemistry, IFR 139, Hôpital Lariboisière, Assistance Publique-Hopitaux de Paris EA 3621, Paris, France; [12]INSERM U955, Faculty of Medicine, Université Paris XII, Créteil, France; [13]FondaMental Foundation, Créteil, France; [14]Saint George's Hospital Medical School, London, UK

Key words: AA-NAT, ADHD, ASMT, genes, melatonin, receptors

Address reprint requests to Thomas Bourgeron, PhD, Human Genetics and Cognitive Functions, Institut Pasteur, 25 rue du Docteur Roux, 75015 Paris, France.
E-mail: thomasb@pasteur.fr
*These authors contributed equally to the work.

Received March 7, 2011;
Accepted April 22, 2011.

Introduction

Attention-deficit/hyperactivity disorder (ADHD) is one of the most common childhood disorders and can continue through adolescence and adulthood. Symptoms include difficulty staying focused and paying attention, difficulty controlling behavior, and hyperactivity [1]. ADHD has three subtypes: predominantly hyperactive-impulsive, predominantly inattentive, and combined hyperactive-impulsive and inattentive. Most children have the combined type of ADHD. The cause of ADHD remains largely unknown, but genetic risk factors are strongly suggested from twin and familial studies [2, 3]. Alteration in sleep is notably frequent in patients with ADHD, and low melatonin levels as well as an abnormal delay of melatonin onset was reported in patients with ADHD and chronic sleep onset insomnia [4–6]. More recently, a maternal impaired serotonin production was proposed as a risk factor for ADHD in the offspring [7]. In addition, the sleep of patients with ADHD is greatly improved by the use of melatonin as a medication [8–12], suggesting that the endogenous level of melatonin is not sufficient to reset the biologic clock. Indeed, Laakso et al. [13] and Leger et al. [14] showed that endogenous melatonin predicts efficacy of exogenous melatonin for sleep disorders [13, 14].

Melatonin is synthesized during the night in the pineal gland and is involved in sleep induction, circadian rhythm regulation, and the immune response, and it functions as an antioxidant [15–18] Melatonin biosynthesis requires serotonin, which is first acetylated by aryl alkylamine *N*-acetyltransferase (AA-NAT) and then converted to melatonin by acetyl serotonin methyl transferase (ASMT also known as hydroxyindole O-methyltransferase or HIOMT) [19]. Melatonin signaling is mediated by the guanine nucleotide binding (G) protein-coupled receptors MTNR1A and MTNR1B that are expressed in the suprachiasmatic nucleus, but are also present in hypothalamic nuclei, retina, and immune cells. One of the downstream cellular effects of melatonin receptor activation is inhibition of adenylate cyclase and cAMP production [20]. GPR50 is an orphan G protein-coupled receptor with no affinity for melatonin, but as a dimer with MTNR1A inhibits melatonin signaling [21].

This study determined whether patients with ADHD could carry an excess of genetic mutations within the melatonin pathway compared with controls and ascertained if these mutations were deleterious at the functional level and finally reported on the clinical profile of the patients carrying the mutations.

Materials and methods

Subjects

Unrelated Swedish participants with DSM-IV-TR ADHD (n = 101, 59 men, 42 women) were recruited at the Sahlgren University Hospital in Göteborg. They have been diagnosed after personal interview, clinical examination and comprehensive testing by an experienced psychiatrist and a neuropsychologist, and after collateral interview with a close family member. The instruments used included SCID-I, SCID-II, ADHD-RS, and DSM-IV-checklists. ADHD had been assigned as 'main diagnosis' in all 101 participants in the ADHD group, but the vast majority had comorbid disorders (e.g. 53% had a history of mood disorder or major depressive disorder 39%; Table S1). Unrelated Swedish participants without ADHD (n = 220, 142 men, 78 women) were recruited through advertisements. They had not undergone extensive neuropsychiatric assessment, but had confirmed, in face to face interviews, that they did not have a diagnosis for ADHD. The presence of sleep problems in controls was not ascertained. All individuals included are of European ancestry. The local research ethics boards reviewed and approved the study. Informed consent was obtained from all participants.

Genetic screening of the melatonin pathway

All PCR and sequencing of *ASMT*, *MTNR1A*/*B*, and *GPR50* were performed as previously described [22, 23]. For *AA-NAT*, amplification of both exons was performed in a single PCR of 1655 bp with primers AA-NAT F: GAATGTGCCCATTGATTTAGG and AA-NAT R – GCCCGGTCTCAGGTACAGAGT. PCR products were sequenced with the BigDye Terminator Cycle Sequencing kit (V3.1; Applied Biosystems, Courtaboeuf, France). Samples were then subjected to electrophoresis, using an ABI PRISM genetic analyzer (Applied Biosystems). For all nonsynonymous mutations, genotyping was confirmed by the sequence of an independent PCR product.

To test whether there was an enrichment of *MTNR1A* and *MTNR1B* mutations in patients compared to controls, each group was divided in three subgroups of individuals carrying 0, 1 or 2 coding variations. Enrichment was tested using a chi-square test. The X-linked *GPR50* gene was analyzed separately by comparing allelic distribution in case and controls after stratification by sex.

Functional analyses of the melatonin pathway

Human cDNA for *AA-NAT* and *ASMT* were cloned in a PcDNA-dest47 vector. Flag-MTNR1A was cloned in pcDNA3 expression vectors as previously described [24]. Mutagenesis was performed with the QuickChange II XL site-directed mutagenesis kit (Invitrogen, Carlsbad, NM, USA). Each clone was purified with Endofree Plasmid Maxi kit (Qiagen, Courtaboeuf, France) and entirely sequenced to rule out additional mutations in the cDNA. Transient transfections in COS cells were performed as previously described [24] using JetPEI (Polyplus Transfection, Illkirch, France) or AMAXA Nucleofector kit (Lonza, Basel, Switzerland). AA-NAT (EC 2.3.1.87) and ASMT (EC 2.1.1.4) activities were measured by radioenzymology as described [23, 25]. The consequence of the splice mutation on the ASMT mRNA was studied in Melke et al., 2008 [23]. RT-PCR showed an abnormal ASMT transcript, encoding a putative truncated ASMT protein, lacking the methyl-transferase domain. The immunofluorescence experiments, immunoblots, binding experiments, and cAMP and ERK1/2 signaling experiments of the MTNR1A Y170X were performed as described [22]. Statistical

differences between wild-type and mutant AA-NAT, ASMT, and MTNR1A activities were tested using a two-tailed *t*-test.

Results

We could detect variations modifying the protein sequence in the five main genes required for both melatonin synthesis (*AA-NAT* and *ASMT*) and signaling (*MTNR1A*, *MTNR1B*, and *GPR50*) (Tables 1 and 2). All coding variations in *GPR50* were frequent poymorphisms. No significant enrichment of coding variations in *MTNR1A* and *MTNR1B* was observed in patients with ADHD compared with controls. Indeed, the proportion of individuals carrying 0, 1, or 2 coding variations was not different in patients with ADHD (70, 26, and 5 subjects respectively) compared with controls (158, 48, and 14 subjects respectively; $\chi^2 = 0.76$, df = 2, $P = 0.68$). Interestingly, one splice site mutation of *ASMT* (IVS5 + 2T > C) and one stop mutation in *MTNR1A* (Y170X) were predicted to cause severe functional alterations and were detected in two independent patients with ADHD, but not in our geographically matched comparison group. The splice site mutation of *ASMT* (IVS5 + 2T > C) was detected in a male patient presenting with ADHD, autistic traits, and sleep disorders (see Table S2 for clinical details). The mutation was inherited from a mother with attention/impulsivity problems and motor control dysfunction. The *MTNR1A* stop mutation Y170X was detected in an adult woman presenting with ADHD and obsessive compulsive disorder

Table 1. AA-NAT, ASMT, MTNR1A, MTNR1B, and GPR50 variants identified in 101 patients with ADHD and 220 control individuals

Variations	Genomic position (bp)	Allele (major/minor)	Number of individual carrying the variation (Allelic frequency %)		Function
			ADHD (N = 101)	Controls (N = 220)	
AA-NAT	(Chr 17)				
V62I	71976870	G/A	1 (0.5%)	1 (0.23%)	As control
A163V	71977511	C/T	1 (0.5%)	3 (0.68%)	Altered
G177D	71977553	G/A	0	1 (0.23%)	Altered
ASMT[a]	(Chr X)				
IVS5 + 2T > C	1708834	T/C	1 (0.5%)	0	Altered
D210G	1712109	A/G	0	1 (0.23%)	Altered
L298F	1721761	C/T	1 (0.5%)	2 (0.46%)	Altered
MTNR1A	(Chr 4)				
G166E	187455399	G/A	2 (1%)	8 (1.8%)	Altered
Y170X	187455386	C/G	1 (0.5%)	0	Altered
A266V	187455099	C/T	4 (2%)	11 (2.5%)	Altered
K334N	187454894	A/T	2 (1%)	2 (0.46%)	Altered
MTNR1B	(Chr 11)				
G24E	92342610	G/A	1 AA 17 AG 83 GG f(A) = 9.4%	0 AA 30 AG 190 GG f(A) = 6.8%	As control
R138C	92354449	C/T	0	2 (0.46%)	Altered
R231H	92354729	G/A	0	5 (1.14%)	Altered
K243R	92354765	A/G	4 (2%)	10 (2.27%)	Altered

ADHD, attention-deficit/hyperactivity disorder.
[a]The nomenclature of *ASMT* variations is based on P46597-1, the protein sequence coding for the functional isoform of ASMT and therefore is different from the paper by Melke et al. For example L326F is now L298F.

Table 2. Variations in *GPR50* identified in 101 patients with ADHD and 220 control individuals

Variations	Genomic position (bp) (Chr X)	Allele (major/minor)	Number of individual carrying the variation (allelic frequency %)	
			ADHD (N = 101)	Controls (N = 220)
S493R				
Male	150349533	G/A	8A/y 51G/y f(A) = 13%	22 A/y 120 G/y f(A) = 15%
Female			2 AA 5 AG 35 GG f(A) = 11%	0AA 13AG 65 GG f(A) = 8%
Del 502–505				
Male	150349569	Ins/del	25 del/y; 34 ins/y f(del) = 42%	65del/y; 77ins/y f(del) = 46%
Female			8 del/del 21 del/ins 13 ins/ins f(del) = 44%	15del/del 35del/ins 28 ins/ins f(del) = 39%
T532A				
Male	150349649	A/G	22G/y 37A/y f(G) = 37%	63 G/y 79 A/y f(G) = 44%
Female			7 GG 22 AG 13 AA f(G) = 43%	15 GG 35 AG 28 AA f(G) = 39%
I606V				
Male	150349871	G/A	28 A/y 31 G/y f(A) = 47%	78A/y 64G/y f(A) = 55%
Female			7 AA 20AG 15 GG f(A) = 40%	10AA 43 AG 25 GG f(A) = 40%

ADHD, attention-deficit/hyperactivity disorder.

(OCD) as well as social impairment and social phobia. The mutation was transmitted from the father without medical or psychiatric disorders. The proband did not transmit the mutation to her son diagnosed for Asperger syndrome.

For the variations identified in patients, measurements of the recombinant enzyme activities indicated that AA-NAT V62I did not change the overall activity whereas A163V caused twofold reduction ($P = 0.08$) (Fig. 1A). Measurements of ASMT activity indicated that the G219X stop mutation – mimicking the splice site mutation (IVS5 + 2T > C) – and the L298F variation strongly affected enzyme activity (G219X: $P = 0.007$ and L298F: $P = 0.008$) (Fig. 1A). These results obtained in vitro are consistent with our previous report of an in vivo ASMT deficiency in cell lines of patients with autism spectrum disorders (ASD) carrying the same variations [23]. Interestingly, the AA-NAT G177D and the ASMT D210G variations detected here in the controls were also shown to alter enzyme activity ($P = 0.004$ and $P = 0.007$).

All mutations of *MTNR1A* and *MTNR1B* receptors were previously detected in patients with ASD or controls [22], except for the *MTNR1A* Y170X stop mutation found only in the ADHD sample. We have previously shown that the majority of these variations alter the receptor functions by preventing cell surface expression and/or modifying pathway-selective signaling properties [22]. The MTNR1A-Y170X mutant migrated as expected at a lower apparent molecular weight (approximately 35 kDa), displayed severely reduced surface expression, and was localized mainly at intracellular membrane compartments (Fig. 1C,D). It was devoid of any binding activity, as well as any capacity of inhibiting forskolin-stimulated cAMP accumulation (Fig. 1E) and promoting ERK1/2 phosphorylation (data not shown).

Discussion

In this mutation screening, we detected coding variations in all genes from the melatonin pathway, most of which affect melatonin synthesis and signaling in vitro. The frequency of rare variations was similar in our sample of Swedish patients with ADHD and controls. These results indicate

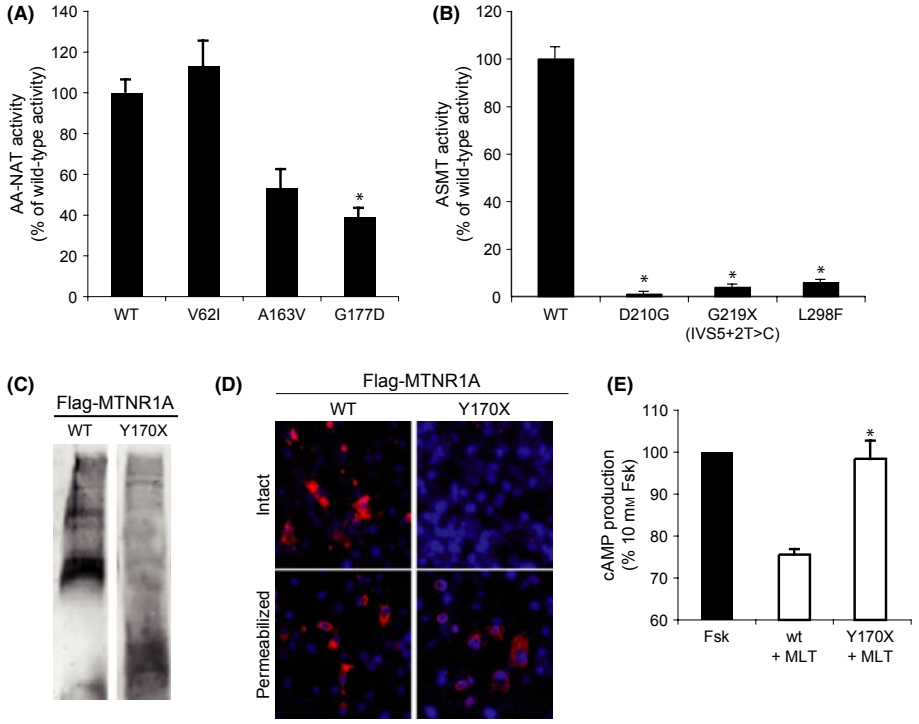


Fig. 1. Functional analyses of the melatonin pathway. (A) AA-NAT activity and (B) ASMT activity, both were measured using COS cells transfected with either the WT or mutant cDNAs. The ASMT G219X mutations removes exons 6–9 from the ASMT sequence and therefore mimics the effect of the IVS5 + T > C splicing mutation observed in the patients with attention-deficit/hyperactivity disorder. The wild-type AA-NAT and ASMT activities are 22 ± 2 nmol/hr/mg prot and 8 ± 1 nmol/hr/mg prot, respectively. (C) Detection of the WT and the MTNR1A-Y170X mutant by SDS–PAGE. Lysates from HEK 293 cells transiently expressing the indicated receptors were separated by SDS–PAGE and analysis performed by Western blot using anti-Flag. (D) Subcellular localization of the wild-type and the MTNR1A-Y170X mutants COS cells transient expression of indicated receptors were permeabilized or not with triton X-100 and total and surface exposed receptors detected by immunofluorescence microcopy with anti-Flag (MTNR1A). Similar results were obtained in three additional experiments. (E) Signaling of MTNR1A mutant through the cAMP pathway. HEK 293 cells were transfected with MTNR1A-Y170X mutant receptor. Inhibition of the cAMP pathway was measured by stimulating cells with forskolin alone (10 μM) or with forskolin and 10 nM melatonin for 60 min. Cyclic AMP levels were determined as described in Materials and methods. Data are means ± S.E.M. of three independent experiments each performed in duplicate (M, *t*-test mutant versus wt $P < 0.01$).

that coding variants of the melatonin pathway are not major risk factors for ADHD. The impact of an alteration in the melatonin pathway is therefore most likely related to a broader phenotype. For example, recent studies reported that *ASMT* variations could be associated with depression [26], but also with ASD [23, 27–29]. The presence of a genetic overlap between ADHD and ASD was recently highlighted by a large-scale twin study in Sweden [30] and by the identification of shared susceptibility genes [31].

In our study, we detected several deleterious mutations of the melatonin pathway in patients with ADHD presenting with autistic features or social impairments. The *MTNR1A* Y170X mutation represents the first stop mutation affecting melatonin receptors in humans and was observed in a patient with ADHD, social impairment, and OCD. In addition, all patients carrying *ASMT* deleterious mutations showed a combination of ADHD, autistic traits, and sleep disorder (Table S2). The same *ASMT* splice site mutation (IVS5 + 2T > C) was previously observed in 0.8% patients with ASD (6/749) and in 0.1% of the controls (1/937; $P = 0.03$) [23, 27, 28]. Melatonin acts on numerous sites in the brain [32] and could modulate synaptic homeostasis of specific neuronal circuits associated with ASD [33]. Interestingly, *MTNR1A* is expressed in the prefrontal cortex and the striatum, which show reduced thickness or volume loss in patients with ADHD [34, 35].

Several limitations are present in our study. First, a larger sample size would allow us to detect other risk alleles of lower frequency, increasing the power of our enrichment test. Second, a more complete phenotypic characterization of the subjects, such as sleep/circadian problems and blood/urinary melatonin levels, would allow us to explore further the consequences of the deleterious mutations. Additional work is therefore needed to replicate the association between abnormal melatonin pathway and increased risk of ADHD symptoms in patients.

In summary, we provide the first genetic and functional ascertainment of defects in the melatonin pathway in patients with ADHD. The impact and the specificity of such melatonin deficiencies remain to be fully characterized. However, owing to the importance of the biologic clock in many physiological functions, the ascertainment of deficits in the melatonin pathway represents relevant information for determining individual risk for circadian disorders.

Acknowledgements

We thank the patients and their families for participating in this study. We thank the Centre d'Investigation Clinique of the Robert Debré Hospital for collecting and processing the biologic samples of this study. This work was supported by the Institut Pasteur, Université Paris Diderot, INSERM, CNRS, FP6 ENI-NET, Fondation FondaMental and Fondation pour la Recherche Médicale ('Equipe FRM', RJ).

Contributions

G.N., H.A., J.M., M.R., O.S., I.C.G., C.G. recruited patients and controls; P.C., E.S., N.L., M.K., G.H., L.B performed mutation screening and P.C., C.D., I.S., F.F., R.T., C.L. performed genetic analyses; N.C., H.G.B, C.P., J.M.L., J.L.G., P.C. carried out biochemical analyses and statistical analyses was carried out by P.C., T.B., R.J., C.P., N.C.; P.C., T.B., R.J., N.C., C.G., R.D., M.L. contributed to conception, design, and manuscript drafting.

Conflict of interest

We certify that there is no conflict of interest regarding the material discussed in the manuscript.

References

1. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, 4th edn. American Psychiatric Press, Washington D.C., 1994.
2. Franke B, Neale BM, Faraone SV. Genome-wide association studies in ADHD. Hum Genet 2009; **126**:13–50.
3. Thapar A, Langley K, Owen MJ et al. Advances in genetic findings on attention deficit hyperactivity disorder. Psychol Med 2007; **37**:1681–1692.
4. Van Der Heijden KB, Smits MG, Van Someren EJ et al. Idiopathic chronic sleep onset insomnia in attention-deficit/hyperactivity disorder: a circadian rhythm sleep disorder. Chronobiol Int 2005; **22**:559–570.
5. Baird A, Siddiqu IA, Forbes-Robertson S et al. Circadian rhythms in adult attention-deficit/hyperactivity disorder. XI Congress of the European Biological Rhythms Society August 2009, Strasbourg, France 2009 (Abstract P12-3).
6. Van Veen MM, Kooij JJ, Boonstra AM et al. Delayed circadian rhythm in adults with attention-deficit/hyperactivity disorder and chronic sleep-onset insomnia. Biol Psychiatry 2010; **67**:1091–1096.
7. Halmoy A, Johansson S, Winge I et al. Attention-deficit/hyperactivity disorder symptoms in offspring of mothers with impaired serotonin production. Arch Gen Psychiatry 2010; **67**:1033–1043.
8. Van Der Heijden KB, Smits MG, Van Someren EJ et al. Effect of melatonin on sleep, behavior, and cognition in ADHD and chronic sleep-onset insomnia. J Am Acad Child Adolesc Psychiatry 2007; **46**:233–241.
9. Weiss MD, Wasdell MB, Bomben MM et al. Sleep hygiene and melatonin treatment for children and adolescents with ADHD and initial insomnia. J Am Acad Child Adolesc Psychiatry 2006; **45**:512–519.
10. Hoebert M, Van Der Heijden KB, Van Geijlswijk IM et al. Long-term follow-up of melatonin treatment in children with ADHD and chronic sleep onset insomnia. J Pineal Res 2009; **47**:1–7.
11. Bendz LM, Scates AC. Melatonin treatment for insomnia in pediatric patients with attention-deficit/hyperactivity disorder. Ann Pharmacother 2010; **44**:185–191.
12. Tjon Pian Gi CV, Broeren JP, Starreveld JS et al. Melatonin for treatment of sleeping disorders in children with attention deficit/hyperactivity disorder: a preliminary open label study. Eur J Pediatr 2003; **162**:554–555.
13. Laakso ML, Lindblom N, Leinonen L et al. Endogenous melatonin predicts efficacy of exogenous melatonin in consolidation of fragmented wrist-activity rhythm of adult patients with developmental brain disorders: a double-blind, placebo-controlled, crossover study. Sleep Med 2007; **8**:222–239.
14. Leger D, Laudon M, Zisapel N. Nocturnal 6-sulfatoxymelatonin excretion in insomnia and its relation to the response to melatonin replacement therapy. Am J Med 2004; **116**:91–95.

15. Reiter RJ, Tan DX, Fuentes-Broto L. Melatonin: a multitasking molecule. Prog Brain Res 2010; **181**:127–151.
16. Jarzynka MJ, Passey DK, Johnson DA et al. Microtubules modulate melatonin receptors involved in phase-shifting circadian activity rhythms: in vitro and in vivo evidence. J Pineal Res 2009; **46**:161–171.
17. Jung KH, Hong SW, Zheng HM et al. Melatonin ameliorates cerulein-induced pancreatitis by the modulation of nuclear erythroid 2-related factor 2 and nuclear factor-kappaB in rats. J Pineal Res 2010; **48**:239–250.
18. Paradies G, Petrosillo G, Paradies V et al. Melatonin, cardiolipin and mitochondrial bioenergetics in health and disease. J Pineal Res 2010; **48**:297–310.
19. Simonneaux V, Ribelayga C. Generation of the melatonin endocrine message in mammals: a review of the complex regulation of melatonin synthesis by norepinephrine, peptides, and other pineal transmitters. Pharmacol Rev 2003; **55**:325–395.
20. Jockers R, Maurice P, Boutin JA et al. Melatonin receptors, heterodimerization, signal transduction and binding sites: what's new? Br J Pharmacol 2008; **154**:1182–1195.
21. Levoye A, Dam J, Ayoub MA et al. The orphan GPR50 receptor specifically inhibits MT1 melatonin receptor function through heterodimerization. EMBO J 2006; **25**:3012–3023.
22. Chaste P, Clement N, Mercati O et al. Identification of pathway-biased and deleterious melatonin receptor mutants in autism spectrum disorders and in the general population. PLoS ONE 2010; **5**:e11495.
23. Melke J, Goubran Botros H, Chaste P et al. Abnormal melatonin synthesis in autism spectrum disorders. Mol Psychiatry 2008; **13**:90–98.
24. Ayoub MA, Couturier C, Lucas-Meunier E et al. Monitoring of ligand-independent dimerization and ligand-induced conformational changes of melatonin receptors in living cells by bioluminescence resonance energy transfer. J Biol Chem 2002; **277**:21522–21528.
25. Chae HD, Park TJ, Lee YK et al. Rapid and simple measurement of serotonin N-acetyltransferase activity by liquid biphasic diffusion assay. Neurochem Int 1999; **35**:447–451.
26. Galecki P, Szemraj J, Bartosz G et al. Single-nucleotide polymorphisms and mRNA expression for melatonin synthesis rate-limiting enzyme in recurrent depressive disorder. J Pineal Res 2010; **48**:311–317.
27. Toma C, Rossi M, Sousa I et al. Is ASMT a susceptibility gene for autism spectrum disorders? A replication study in European populations. Mol Psychiatry 2007; **12**:977–979.
28. Jonsson L, Ljunggren E, Bremer A et al. Mutation screening of melatonin-related genes in patients with autism spectrum disorders. BMC Med Genomics 2010; **3**:10.
29. Cai G, Edelmann L, Goldsmith JE et al. Multiplex ligation-dependent probe amplification for genetic screening in autism spectrum disorders: efficient identification of known microduplications and identification of a novel microduplication in ASMT. BMC Med Genomics 2008; **1**:50.
30. Lichtenstein P, Carlstrom E, Rastam M et al. The genetics of autism spectrum disorders and related neuropsychiatric disorders in childhood. Am J Psychiatry 2010; **AiA**:1–7.
31. Williams NM, Zaharieva I, Martin A et al. Rare chromosomal deletions and duplications in attention-deficit hyperactivity disorder: a genome-wide analysis. Lancet 2010; **376**:1401–1408.
32. Maronde E, Stehle JH. The mammalian pineal gland: known facts, unknown facets. Trends Endocrinol Metab 2007; **18**:142–149.
33. Toro R, Konyukh M, Delorme R et al. Key role for gene dosage and synaptic homeostasis in autism spectrum disorders. Trends Genet 2010; **26**:363–372.
34. Konrad K, Eickhoff SB. Is the ADHD brain wired differently? A review on structural and functional connectivity in attention deficit hyperactivity disorder. Hum Brain Mapp 2010; **31**:904–916.
35. Soliva JC, Carmona S, Fauquet J et al. Neurobiological substrates of social cognition impairment in attention-deficit hyperactivity disorder: gathering insights from seven structural and functional magnetic resonance imaging studies. Ann N Y Acad Sci 2009; **1167**:212–220.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Clinical description of the ADHD sample.

Table S2. Clinical observations of patients with ADHD carrying rare genetic variations of the melatonin pathway.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.